A Systematic Literature Review of Recent Advances on Context-Aware Recommender Systems

Pablo Mateos[1] and Alejandro Bellogín[1*]

[1*]Computer Science Department, Universidad Autónoma de Madrid, Madrid, 28049, Spain.

*Corresponding author(s). E-mail(s): alejandro.bellogin@uam.es; Contributing authors: pablo.mateosm@estudiante.uam.es;

Abstract

Recommender systems are software mechanisms whose usage is to offer suggestions for different types of entities like products, services, or contacts that could be useful or interesting for a specific user. Other ways have been explored in the field to enhance the power of these systems by integrating the context as an additional attribute. This inclusion tries to extract the user preferences more accurately taking into account multiple components such as temporal, spatial, or social ones. Notwithstanding the magnitude of context-awareness in this area, the research community is in agreement with the lack of framework for context information and how to integrate it into recommender systems. Under this premise, this paper focuses on a comprehensive systematic literature review of the state-of-the-art recommendation techniques and their characteristics to benefit from contextual information. The following survey presents the following contributions as outcomes of our study: i) determine a framework where multiple aspects are taken into account to have a clear definition of context representation, ii) the techniques used to incorporate context, and iii) the evaluation of these methods in terms of reproducibility and effectiveness. Our review also covers some crucial topics about context integration, classification of the contexts, application domains, and evaluation of the used datasets, metrics, and code implementations, where we observed clear shiftings in algorithmic and evaluation trends towards Neural Network approaches and ranking metrics, respectively. Just as importantly, future research opportunities and directions are exposed as final closure, standing out the exploitation of various data sources and the scalability and customization of existing solutions.

Keywords: Recommendation systems, Context awareness, Context modeling, Evaluation

1 Introduction

Recommender systems (RS) present information to users about a series of items based on their preferences, and try to predict which item is more appropriate to these users. As an inherent part of this task, these systems foster the idea of suggesting relevant items in a vast and unknown space where the user would be incapable to explore. Therefore, to offer this open variety to users, recommender systems have the goal of learning user selections to set relationship patterns. To establish these connections, this information could be collected explicitly, by means of user ratings, or implicitly through user behavior.

Recommender systems are traditionally classified according to the type of information they use. This gives rise to two large groups: those based on Collaborative Filtering (CF) [1] and those based on content (CB) [2]. Hybrid algorithms are also often considered as a third type, since collaborative information is usually combined with content to avoid the problems of each of these technologies [3]. More specifically, CF recommender systems exploit the underlying patterns of user preferences that are acquired from users to create personalized recommendations [1]. On the other hand, CB recommendation systems analyze the content of items the user has liked in the past [2]. However, recommender systems do not rely solely on such information to generate recommendations, but can also rely on other data sources depending on the use case.

For that matter, the concept of **context** came into play. Context is a multifaceted concept that has been used across different disciplines. The most cited definition of context in recommender systems is given by Abowd and Dey in [4], who defined it as '*any information that can be used to characterize the situation of an entity, where an entity can be a person, place, physical, or computational object*', even though this work was not specifically oriented to recommendation, but to human-computer interaction. This additional information can be temporary, geographical, or any other type that can help the system to be more precise in its objective, and those systems who exploit it are called *context-aware recommender systems* (CARS).

Usually, recommender systems take into account two variables: users and items, but as mentioned before, a new dimension can coexist, which we will call "contextual information". For example, information about the weather can help a tourist company to improve its recommender system. In this scenario, it has been shown that the incorporation of this type of information makes it possible to better characterize and profile users using more detailed data and, therefore, make better recommendations. Traditional approaches, such as classic CF methods, are not suitable for these situations, unless adapted appropriately. Indeed, another advantage of exploiting contextual information is to alleviate the problem of missing information or very sparse preferences. Moreover, users make their decisions under certain contextual circumstances, so taking relevant contextual information into account when making recommendations helps to obtain more accurate predictions of user preferences [5, 6].

Throughout 2000s, and in particular in 2010, some approaches were introduced, mostly focused on rating prediction; then, formalizations of the problem and the first classifications of the approaches emerged. Starting with the seminal work of Adomavicius et al. [7], that work discusses the main challenges and illustrates the usage of this type of algorithms, focusing on travel and music domains, although foreseeing more interactive or adaptive scenarios like conversational applications. More recently, we find reviews specialized in particular domains, like social networks [8], cultural heritage [9], or temporal aspects [10]. Additionally, some of these reviews have been focused on the algorithmic part [11], the applications [12], or its challenges and opportunities [8]. In contrast, this survey aims to present the (recent) literature in a systematic, unbiased way, without focusing in any specific domain, while capturing all the different dimensions of these works (context types, algorithms, evaluation, and so on). The survey closest to ours would be the work presented in [13], where the authors also conducted a systematic literature review, but focused mostly on classical CB, CF, and hybrid techniques. Our review covers more aspects and, by being more recent, it analyzes a wider range of algorithm techniques. On top of that, a unique contribution that stems from our study is its focus on reproducibility, by analyzing the extent of existing public repositories or open source implementations.

Despite its increasing use in the area, there is no consensus around the definition of context or across which characteristics could be categorized upon. In this work, we consider three representative aspects for context information: acquisition, observability, and dynamism of its context characterization, as shown in previous studies [9, 11, 14]. The first term refers to whether the context can be previously acquired and defined or if it is tied to user interactions; the second aspect is linked to how explicit the context can be provided; finally, the last aspect is related to whether the contextual values change over time. Besides that, we have also introduced a classification in five different categories to segment the context characterization following the schema exposed in [15]. Beyond these characteristics, we shall also categorize the literature according to the domains where the context is exploited (typically e-commerce, tourism, multimedia, etc.) and, from an algorithmic point of view, which recommendation approaches are more frequently considered when integrating contextual information, and how this information is processed and evaluated in CARS.

We considered a total number of 90 studies, published from 2014 to 2024, to perform this systematic literature review. To the best of our knowledge, this is the most up-to-date review conducted in the field of context-aware recommender systems that follows a standardized and repeatable protocol. We expect our study could be useful to other researchers working in this area, especially for better identifying possible approaches and future trends.

The remainder of this work is structured thus: in Sect. 2, we detail the protocol that we followed for conducting the review. Then, we present the quantitative results and main findings in Sect. 3, and we provide a possible interpretation of the outcomes of the review in Sect. 4. Finally, we conclude our work with Sect. 5.

2 Methodology

We decided to perform this review according to the guidelines designed by Kitchenham and Charters for Systematic Literature Reviews (SLR) in the field of Software Engineering [16]. This method guarantees that the outcome of the review is verifiable and repeatable by other researchers.

In the following subsections, we present the research questions addressed in this survey, how the works were selected, and an objective way to assess the quality of the selected studies.

2.1 Research questions

The purpose of this systematic literature review is to identify the studies describing context-aware recommender systems, while understanding the types of contexts being used, the techniques employed, the experimental protocols used to validate them and their reproducibility level, and the most promising directions for the future. For these reasons, we defined the following research questions and address them throughout the paper.

RQ1 What are the most relevant studies addressing context-aware RSs? (Section 3.1)
RQ2 What is the current state-of-the-art regarding context-aware recommenders? In particular:

- RQ2.1 What type of context is exploited in the literature? (Section 3.2)
- RQ2.2 What are the recommendation approaches used by/in context-aware recommenders? (Section 3.3)
- RQ2.3 Which contextual techniques and methods have been proposed for context-aware recommendation? (Section 3.4)
- RQ2.4 In which domains context-aware recommender systems are applied? (Section 3.5)

RQ3 How context-aware recommenders are evaluated in literature? More specifically:

- RQ3.1 Are improvements always achieved when incorporating contextual dimensions? In particular, which baselines are considered when CARS are evaluated? (Section 3.6)
- RQ3.2 Which metrics are considered during their evaluation? (Section 3.7)
- RQ3.3 Which datasets are used for testing the algorithms? (Section 3.8)
- RQ3.4 How reproducible, in principle, state-of-the-art methods are? What is the reproducibility level of the methods evaluated in the literature? (Section 3.9)

RQ4 What are the most promising directions for future works? (Section 3.10)

2.2 Selection process

To retrieve the papers related to context-aware recommendation systems, we have defined a combination of filters and keywords to conduct a bibliographic search of conference proceedings and journal papers in different scientific digital libraries, in particular, Scopus and DBLP. These sources have been selected as reference due to the quality and relevance of their publications. We have used the following advanced search strings, one specifically for journals and another one for conferences. The most recent execution of these two queries took place in April 2024.

The first query corresponds to paper selection from journals:

```
(TITLE ( "context-aware" OR "contextual" AND
"recommendation systems"
OR "recommendation" OR "recommender system" )
OR ABS ( "context-aware" AND "recommendation systems"
OR "recommender systems" )
AND NOT TITLE ( "empirical" OR "comparison" OR "experimental"
```

```
OR "survey" OR "evaluation" OR "comparative" ) )
AND PUBYEAR > 2013 AND PUBYEAR < 2025 AND
( LIMIT-TO ( SUBJAREA , "comp" ) ) AND
( LIMIT-TO ( LANGUAGE , "english" ) )
AND ( LIMIT-TO ( DOCTYPE , "ar" ))
```

The second query corresponds to paper selection from conferences, however, due to the length of this second query (owing to the inclusion of a detailed list of specific conferences at the end), we do not include the full query, but its main structure:

```
(TITLE ( "context-aware" OR "contextual" AND
"recommendation systems" OR "recommendation"
OR "recommender system" ) OR ABS ( "context-aware"
AND "recommendation systems" OR "recommender
systems" )
AND NOT TITLE ( "empirical" OR "comparison" OR "experimental"
OR "survey" OR "evaluation" OR "comparative" ) )
AND PUBYEAR > 2013 AND PUBYEAR < 2025 AND
( LIMIT-TO ( SUBJAREA , "comp" ) )
AND ( LIMIT-TO ( LANGUAGE , "english" ))
AND ( LIMIT-TO ( DOCTYPE , "cp" ))
```

We now show which conferences were marked to be considered in this query, where we valued those highly ranked in international conference rankings, such as CORE[1] and which, consistently, appeared in our literature review:

- AAAI Conference on Artificial Intelligence (AAAI)
- ACM Conference on Human Factors in Computing Systems (CHI)
- ACM Conference on Information and Knowledge Management (CIKM)
- ACM International Conference Proceeding Series (ICPS)
- ACM Recommender Systems Conference (RecSys)
- ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD)
- Advances in Intelligent Systems and Computing Conference (AISC)
- Annual Conference of the Association for Computing Machinery Special Interest Group in Information Retrieval (SIGIR)
- Annual Conference on Neural Information Processing Systems (NIPS)
- CEUR Workshop Proceedings
- International Conference on User Modeling, Adaptation, and Personalization (UMAP)
- International Conference on Web Intelligence (WI)
- International Florida Artificial Intelligence Research Society Conference (FLAIRS)
- International Joint Conference on Artificial Intelligence (IJCAI)
- Lecture Notes in Computer Science (LNCS)

As it can be observed, we have applied various filter criteria to select the papers included in the study: i) publication date, papers selected between 2014 and 2024; ii) publication type, subject area and language, where we excluded surveys, comparative studies, workshop and symposium proceedings, computer science related, and non-English papers; iii) in terms of relevance, we have excluded those with publication impact indexes ranking below the second quartile; and iv) more than 10 citations in conference papers published before than 2022.

---

[1] https://portal.core.edu.au/conf-ranks/

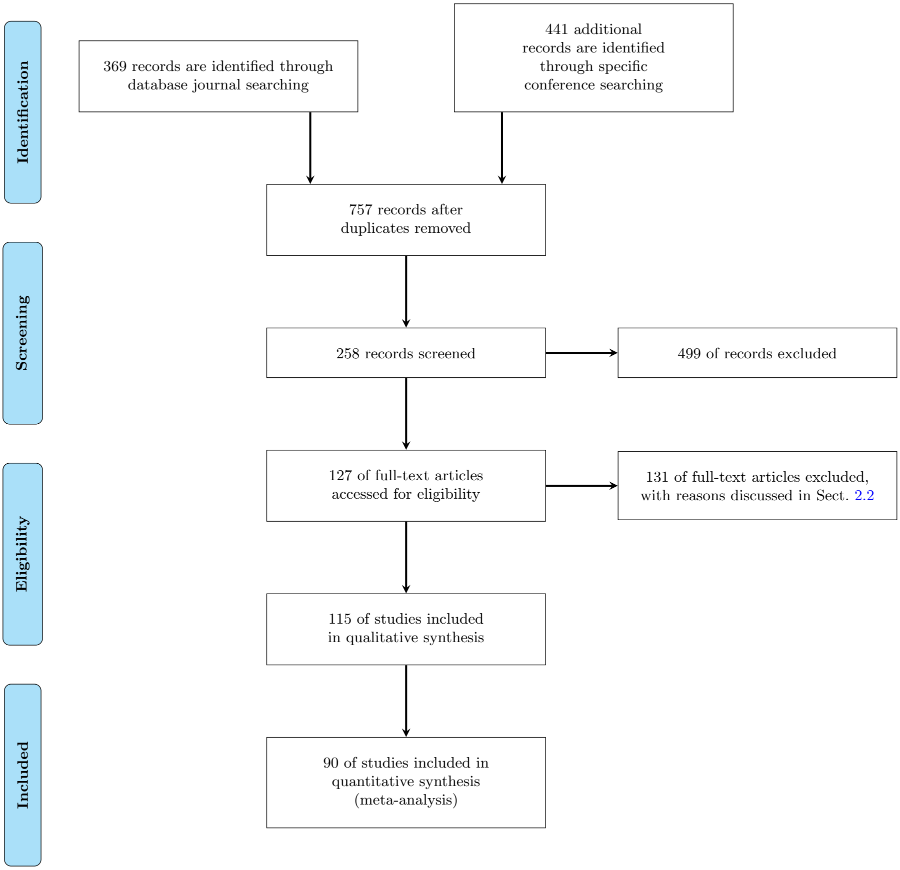

Fig. 1 A PRISMA flow diagram illustrating the full selection process.

After this third filter, we obtained a total of 258 articles. Last but not least, some of the most important conferences in Recommendation Systems, as aforementioned, have been explicitly reviewed to collect papers not indexed in the primary query, such as those presented in workshops at these conferences.

Once the initial collection has been established, we have analyzed carefully and thoroughly all these 258 articles and identified them according to seven criteria: i) type of contribution: framework, method, survey, or metric; ii) recommendation techniques: algorithm and its corresponding family used in the recommendation pipeline; iii) context types and context characteristics, such as observability, acquisition, dynamism, characterization and domain application; iv) paradigm for incorporating context, whether it is pre-filtering, post-filtering, or contextual modeling; v) evaluation: the methods and metrics used to validate the performance of the proposal; vi) data sets (when reported): the data used to train and evaluate the solution; and, finally, vii) the implementation (when reported): the source code used to reproduce the experiments and results outlined in the articles.

After applying the latter filters, we have finalized with 90 curated papers for the ultimate analysis. Figure 1 details and summarizes the full selection process.

Table 1 The quality questions.

Code	Quality question
QQ1	Did the study clearly describe the context?
QQ2	Did the study review the related work?
QQ3	Did the study compare with other approaches?
QQ4	Did the study describe the components of the system?
QQ5	Did the study compare the results against non context-aware baselines?
QQ6	Did the study apply the solution over real scenario?
QQ7	Did the study apply the solution over multiple domains?
QQ8	Did the study include an open source implementation?
QQ9	Did the study review any future lines of work?

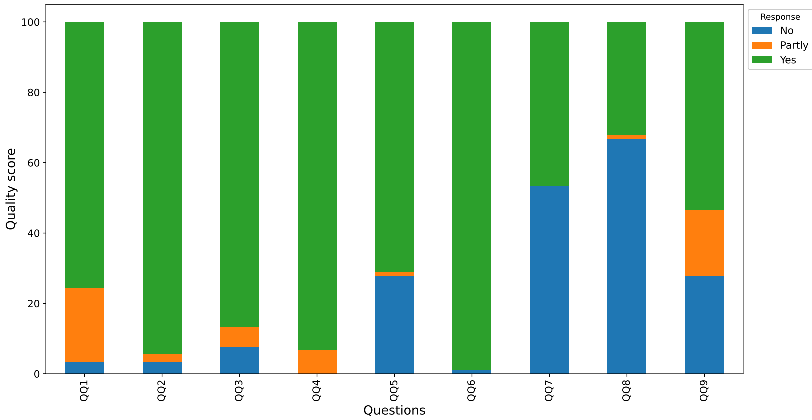

Fig. 2 Responses obtained for the quality questions.

2.3 Quality assessment

To assess the quality of our selection, we have defined a series of quality checks as part of our review (see Table 1). In this sense, we have assigned to each question a certain score depending on the level of completeness of the question when applied to each selected study. Therefore, the possible outcomes, Yes, Partly, or No, have been mapped to numerical values 1, 0.5, or 0.

For this quality assessment, we have assigned an answer for all the selected papers included in the review. Every paper has been scored by both authors independently. If any disagreement or doubt arose in the score, thereupon it was discussed to reach a common resolution. As an ultimate result, we finally computed a global quality number by averaging all the scores, either from the corresponding questions (to get a score in a paper basis) or from the papers (to get a score per quality question, see Figure 2).

3 Results

In this section, we present the findings of our systematic literature review regarding context-aware recommender systems, according to the research questions introduced

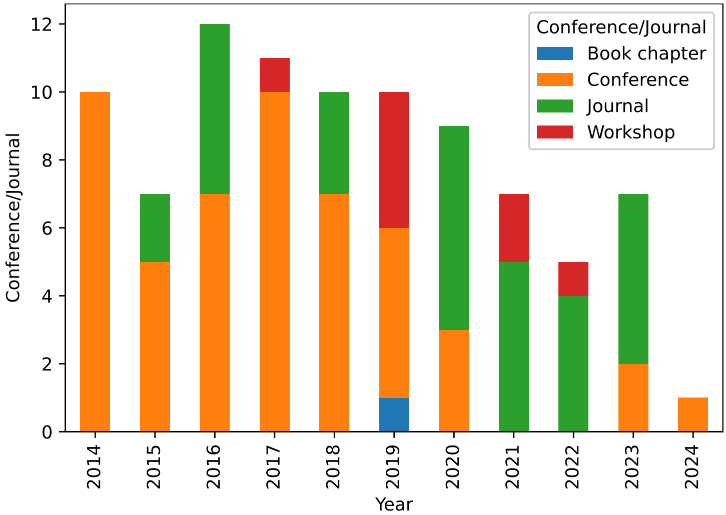

Fig. 3 This stacked barplot represents the number of included studies per year of publication.

in Section 2.1. Hence, we show and summarize the results, and define any necessary concept needed to understand such results. Later, in Section 4, we will discuss these results to produce conclusions that will be reported at the end of the work.

3.1 Included studies

The main purpose of **RQ1** is to identify the studies related to the topic of context-aware recommender systems to be included in this review. Following the protocol detailed in Section 2.2, we identified a total number of 258 studies. These works have been presented mainly in conferences and journals, as shown in Figure 3, although few works have also been published as book chapters or workshop papers. From this figure, we also observe that the period 2016-2019 was the most productive one[2].

Moreover, we applied the quality questions described in Section 2.3 to all the papers included in this survey. Figure 2 shows the average quality score for each quality question. It is possible to observe that the highest score (or the question with more positive answers) is QQ4 (*Did the study describe the components of the system?*), whereas the lowest score is associated to QQ8 (*Did the study include an open source implementation?*). QQ1 (*Did the study clearly describe the context?*) and QQ6 (*Did the study apply the solution over real scenario?*) are tied with the least number of negative responses, after QQ4.

At the same time, in Figure 4, we report the quality scores according to the publication venue. Somewhat surprisingly, journal papers obtained the same score (in average) with respect to conference papers; however, this is a signal that the filters we did on conferences were focused on importance and adherence to the survey topic.

[2]We should note that, because of their relevance, the following 3 papers (previous to 2014) will be mentioned throughout our discussion, although they were not included in our analysis of the literature: [17–19].

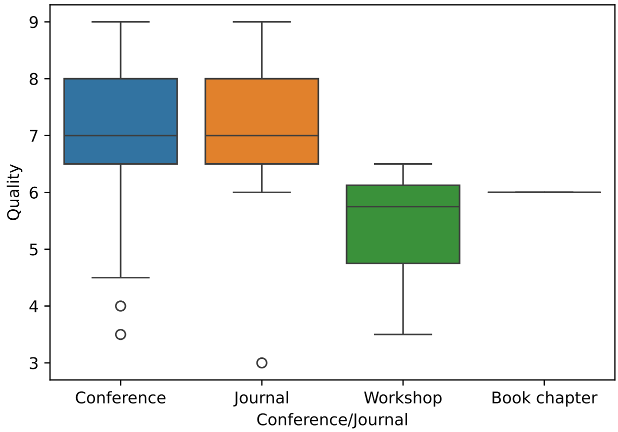

Fig. 4 This boxplot represents the distribution of quality scores per publication type.

3.2 Types of context

As mentioned in the introduction, we have categorized the concept of context in CARS into three aspects: acquisition, observability, and dynamism. Additionally, a characterization classification has been employed to further segment it into five distinct categories. In the subsequent discussion, we will analyze these fundamental characteristics that are clearly and inherently linked to the type of context used in the context-aware recommender systems literature, thereby providing an answer to **RQ2.1**.

3.2.1 Acquisition

The acquisition aspect can be classified into two distinct approaches: representational and interactional. The representational approach conventionally perceives context as being accessible through a predefined set of observable static attributes, where all potential values are fixed and known. Consequently, these attributes remain unchanged over time. Additionally, this approach assumes that the system explicitly observes contextual factors that influence user behavior. An illustrative example of the representational view of context can be found in [20]. The authors treat context as something that can be accessed or obtained. They illustrate this by including contextual factors as additional features in a graph-based search problem to provide recommendations for a specific user. Furthermore, the authors introduce an additional step to manage the context factors in the query, using a post-filtering strategy. In these cases, sensors from different nature (such as physical or hardware, virtual or software, social, and human) may capture the relevant contextual factor applicable in a specific moment. For example, in [21] smart mobile sensors are used to collect data and infer user contexts in an unsupervised way; more general approaches were considered in [22], where sensors from the Internet of Things allow to deal with sparse data.

On the contrary, the interactional perspective characterizes context as a relational entity that is interconnected between activities and dynamically defined. This is illustrated in Figure 5, where the interconnections among the three types of context are depicted, and specially interactional and dynamic aspects work closely together.

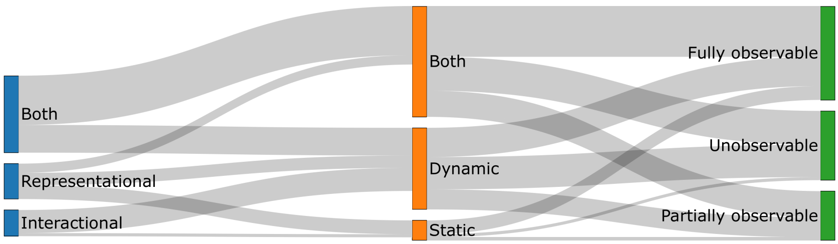

Fig. 5 A Sankey diagram illustrating the interrelationships among various types of context (acquisition on the left part, observability on the right, and dynamism at the center).

In this view, contextual attributes may evolve over time, and it is not possible to determine how many of them will emerge before an activity commences. Furthermore, this approach acknowledges that a contextual factor may only be relevant in relation to a particular activity. A clear example of this concept is presented in the study by Hidasi and Tikk [23], where context is considered as an event, implying that its value is not solely determined by the user or the item, but rather tied to the transaction itself. As a result, context, in this case, is defined by factors such as seasonality and sequential state, which play a significant role in shaping the transaction.

While the representational approach is typically the default accessibility state for context, there are instances where certain factors, particularly those related to the user's psychological features like mood or intent, are not easily accessible to be represented. Complicating matters further, context is sometimes revealed only after the activity or action has been completed, rather than beforehand or independently of its occurrence. This ambiguous nature creates a gray area, leading to an unclear definition of this aspect, and it becomes plausible for both views to coexist harmoniously. To exemplify this, we discuss the work conducted by Zheng et al. [24], where a Hierarchical Collaborative Embedding model is introduced. This model effectively leverages both users' implicit feedback (interactional) and contextual information (representational) derived from the text. By incorporating these aspects, the model jointly learns embeddings of users and items, leading to improved performance in recommendation tasks.

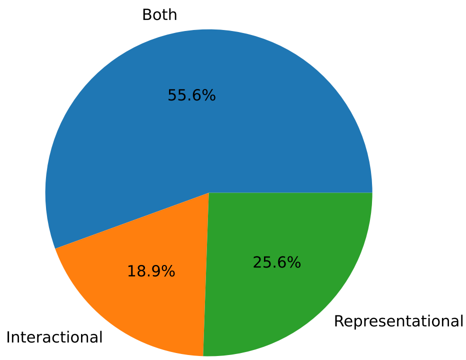

Fig. 6 Context types according to acquisition.

As depicted in Figure 6, the distribution of this aspect is as follows: the most prevalent scenario involves a combination of the two views (Both), followed by the Representational approach standing alone, and finally, the least represented view is the Interactional approach.

3.2.2 Observability

Another classification of context, which extends beyond acquisition, is based on observability, as demonstrated in the work by Adomavicius [7]. This classification encompasses three categories: fully observable, partially observable, and unobservable contexts. In the fully observable context, all contextual factors are explicitly known, aligning closely with the representational view, as illustrated in Figure 5. In contrast, the partially observable and unobservable contexts arise when information is either partially available or not explicitly provided, requiring the use of latent variables for modeling. It is also worth noting, that there is a thin line between those two states and therefore there are different levels of "partial observability" [7].

For instance, the time-aware Recommender Systems introduced in [25] fall into the category of partially observable context. These RS models are designed to capture the evolution of users and items over time, accounting for a partially observable context that is expected to influence users and items and subsequently affect their behavior. Despite having access only to users' behaviors explicitly, the RS effectively models the dynamics occurring due to underlying hidden context structure.

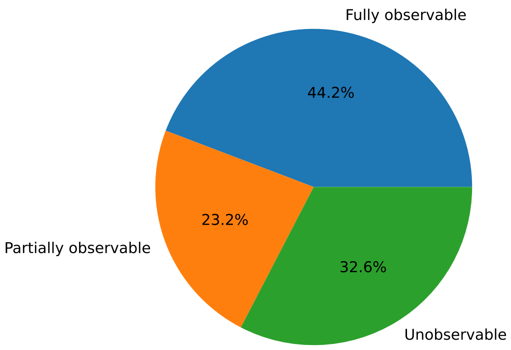

Fig. 7 Context types according to observability.

The majority of studies (about 44.2%), as observed in Figure 7, feature a fully observable context, with the representational view being prevalent in both the Fully Observable and Both scenarios. The second most common scenario is the unobservable context, while the partially observable context, which is the most challenging to handle, constitutes the least represented category.

3.2.3 Dynamism

Another taxonomy to characterize context involves the feature of dynamism. This aspect is categorized into two groups: Static, where contextual factors and their structure remain unchanged over time, and Dynamic, where contextual factors undergo changes. Dynamic factors are more intricate than static ones as they involve evolving data, which makes their evaluation challenging.

As other context types, both scenarios can coexist, especially when contextual information is derived from a combination of both static and dynamic worlds. Based on our literature review, shown in Figure 8, we noticed that over half of the studies incorporate dynamic contextual information, which is likely due to the pervasive nature of time as a crucial factor in various situations. On the other hand, the remaining studies fall under the Static category, where factors like gender, location, or item categories are predetermined and constant.

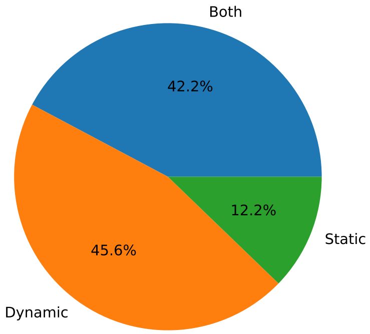

Fig. 8 Context types according to dynamism.

Part of the reason why time and other dynamic contextual attributes can be analyzed is thanks to Mobile Computing and the importance of user mobility in the field [26, 27]. Advances in those areas make easier to capture data – as already mentioned in previous Section 3.2.1 – and to present contextualized recommendations to users [28].

Lastly, we want to emphasize that 42.2% of the studies combine both dynamic and static scenarios, coexisting in harmony. In the article cited as [29], a Graph Convolution Machine is introduced. It incorporates an encoder, graph convolution layers, and a decoder, capable of handling both static and dynamic contexts. The model is evaluated using a book reviews dataset, where static item features include attributes like price and brand, while each review record includes dynamic features such as year, month, day, day of the week, and the user's last purchase.

3.2.4 Context characterization

In this systematic literature review, we follow Villegas & Müller's work [15] regarding their classification of context information. In that work, the authors classify the context that can be used in CARS into five broad categories: individual, location, time, activity, and relational. This characterization can be summarized as follows:

- *Individual context*: it corresponds to information observed from independent entities that may share unique characteristics. This category can be classified into natural, human, artificial, or groups of entities. Human context describes user behavior and preferences (e.g., user payment preferences). Artificial context refers to entities resulting from human actions or technical processes (e.g., hardware and

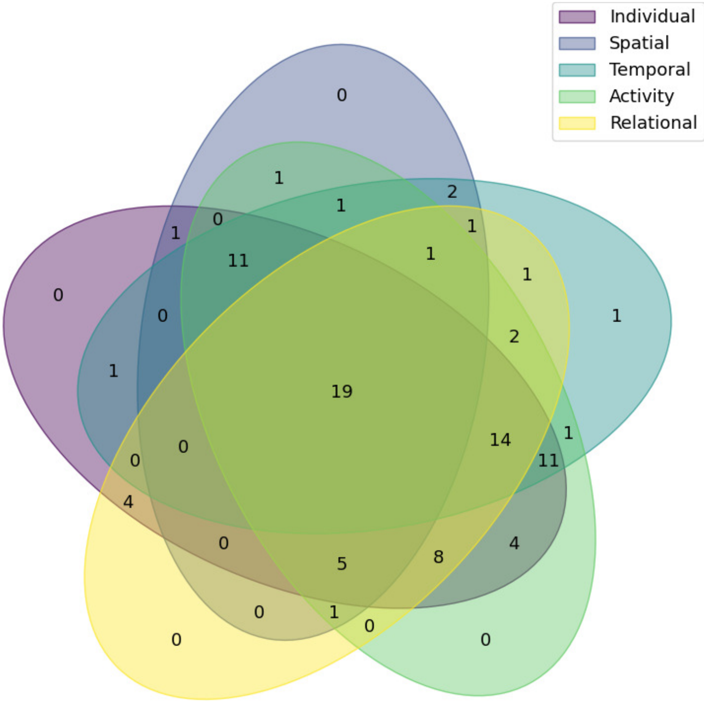

Fig. 9 Context types according to their characterization.

software configurations used in commerce platforms). The last subcategory, groups of entities, refers to groups of independent subjects that share characteristics and can relate to each other (e.g., user preferences in the user's social network).

- *Spatial context*: contextual information associated with the location of the entity. Virtual scenarios like information containing digital coordinates such as the IP address are also contemplated. Location-aware recommender systems are, in fact, a typical application of this type of context, including domains like tourism or cultural heritage [30, 31].
- *Time context*: it corresponds to information such as time of day, current time, week day, and season of the year. Temporal context can be classified as defined or undefined. Defined context indicates time frames with specific start and end points. Undefined context refers to recurring events that occur while another situation is happening, thus not having a specific duration.
- *Activity context*: information related to the activities related by the entities of the system.
- *Relational context*: it refers to the relationships between entities that arise from the circumstances in which the entities are involved. Relational context can be defined as social (i.e., interpersonal relationships such as associations or affiliations) or functional (i.e., the use one entity makes of another).

From our examination of the studies (summarized in Figure 9), it is evident that a majority of them focus on employing Individual and Activity context elements (4 in isolation, 11 with Temporal context, and 14 with Relational context). These choices are particularly pertinent since they closely relate to the fundamental components of the systems, such as items and users, and the preferences intentionally designated by the users themselves. The other context types can be categorized into three groups, with Time being the most prominent among them. This prominence is due to the fact that context is frequently constrained within specific time periods.

3.3 Recommendation approaches

In this section, our focus will be on context-aware recommender systems and the approaches used to address the various challenges they face. These approaches will provide insights into answering **RQ2.2**. We have identified four distinct families of approaches commonly used in the literature, as summarized in Figure 10 and broken down by algorithm methodology in Table 2. It is important to note that a single study may use multiple approaches. Moreover, in Figure 11 we show the evolution of these approaches throughout the years, evidencing the recent popularity of studies exploiting Neural networks.

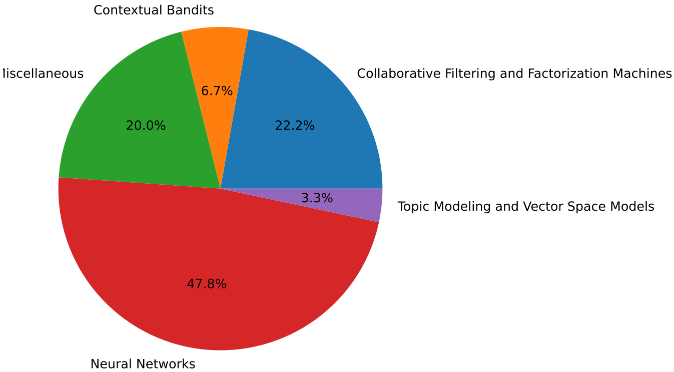

Fig. 10 Percentage of studies where each type of recommendation approach was used.

3.3.1 Topic Modeling and Vector Space Models

In the domain of CARS, advanced techniques like topic modeling and vector space models play a crucial role in enhancing the accuracy and relevance of recommendations. These methodologies enable the system to extract meaningful patterns and relationships from the vast amount of available data, thereby providing users with more contextually relevant suggestions.

On one hand, topic modeling is a text analysis technique that aims to discover latent semantic structures within a collection of documents. By identifying underlying topics and their associated keywords, topic modeling facilitates a deeper understanding of the content without relying solely on explicit keyword matches. In context-aware recommendation systems, topic modeling can be employed to uncover the intrinsic themes within user-generated content, such as reviews, comments, or textual interactions. One widely used topic modeling technique is Latent Dirichlet Allocation (LDA). LDA assumes that each document is a mixture of various topics

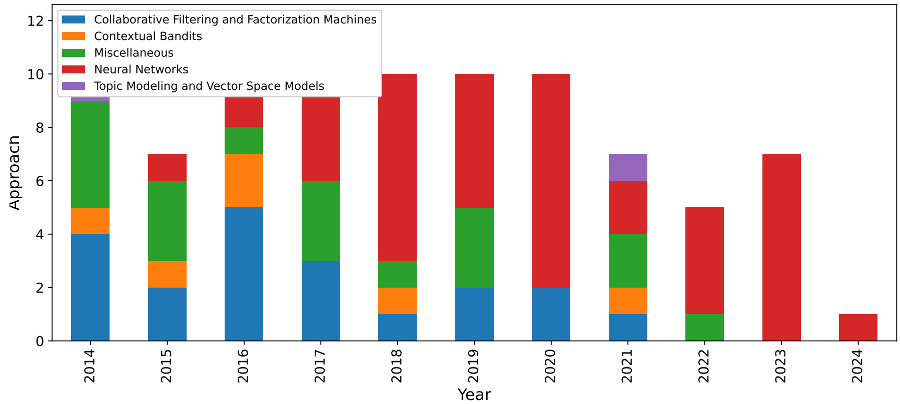

Fig. 11 Number of studies throughout the years when each type of recommendation approach was used.

and that each topic is characterized by a distribution of words. For example, in [32], they address the context-aware problem by extracting users' interests from their digital traces with a novel unsupervised approach called Channel-Aware Latent Dirichlet Allocation (CA-LDA), identifying topics that users frequently are more engaged with.

On the other hand, vector space models represent textual information as numerical vectors in a multi-dimensional space. These models capture the semantic relationships between words and documents, enabling efficient similarity calculations and information retrieval. In the context of a recommendation system, vector space models offer a way to quantify the relationships between users, items, and contextual factors. In this sense, we have the following example from [19], where the authors propose an enhanced Vector Space Model (eVSM) framework to incorporate contextual details, employing distributional semantics. This method creates a contextually sensitive user profile by combining user preferences with a semantic vector representation of the context, allowing for more personalized and relevant recommendations.

These methods have been recognized on 3 occasions within the overall count of analyzed studies.

3.3.2 Collaborative Filtering and Factorization Machines

Collaborative filtering methods, like matrix factorization and nearest neighbors, are fundamental techniques that play a pivotal role in context-aware recommendation systems. These methodologies enable the system to leverage user-item interactions and uncover latent patterns within the data. Classical Collaborative filtering algorithms (i.e., nearest neighbors) can be categorized into two types: user-based and item-based. The user-based collaborative filtering algorithm suggests items that similar users have shown interest in, whereas the item-based algorithm recommends items that are similar to those previously interacted with by the user. Given the number of items typically is significantly larger than users, many recommendation systems opt for the item-based collaborative approach over the user-based one, primarily due to considerations related to computational complexity.

The most significant contrast between the Matrix Factorization (MF) and classical CF algorithms lies in their operational approach: MF involves a learning process,

Table 2 Algorithm methodologies broken down by families of approaches.

Family Approach	Algorithm Methodology	Used in
Topic Modeling and Vector Space Models	Context-Aware LDA	[32–34]
	Context-Aware Word-Based	[34]
Collaborative Filtering and Factorization Machines	Context-Similarity Sparse Linear Method - Matrix Factorization	[35]
	Fast Factorization Machine	[36]
	Gaussian Process Factorization Machines	[37]
	Gradient Boosting Factorization Machines	[38]
	Item-based Collaborative Filtering + Semantic Similarity and LoD	[39]
	Lambda Matrix Factorization	[40]
	Matrix Factorization + Linear Model	[23, 41–43]
	Matrix Factorization + Deviation-based Pro-Filtering	[44]
	SVD + Pre and Post-Filtering	[45]
	Tensor Factorization (TF)	[46–48]
	TF + Adversarial learning	[49]
	TF + Bayesian Personalized Ranking	[50]
	TF + Collaborative Filtering	[51]
	TF + Hierarchical Bayesian Method	[52]
	TF + Label Ranking	[53]
	User-based Collaborative Filtering	[54]
Contextual Bandits	Collaborative Contextual Linear Bandits	[55]
	Constrained Contextual Bandit	[56]
	Contextual Multi Armed Bandits	[57]
	Ensemble Contextual Bandits	[58]
	Parameter Free Contextual Bandits	[59]
	Time Varying Multi-Armed bandit	[60]
Neural Networks	Autoencoders	[21, 61–64]
	BERT encoder + Graph Convolutional Network + GPT2	[65]
	BERT plus Embedding Layer Generation	[66]
	Consesus Clustering + Deep Learning Collaborative Filtering	[67]
	Convolutional Factorization Machine	[68]
	Embedding-based Neural Networks	[69–71]
	Fined-tuned GPT-4	[72]
	Graph Neural Network	[29, 73]
	Hierarchical Collaborative Embedding	[24]
	Neural Collaborative Filtering	[74–76]
	Neural Network + Attention Layer	[77–84]
	Neural Network + Factorization Matrix	[85, 86]
	Neural Network + Prefiltering	[87]
	Neural Network + Variational Bayesian Approach	[88]
	Neural Probabilistic Model	[89]
	Recurrent Neural Networks	[6, 90–93]
	Transformer-based encoder Network	[94]
Miscellaneous	Boosting Methods	[95, 96]
	Content-based filtering plus Semantic and Entity Linking	[97]
	Contextual Ensemble	[98]
	Contextual Heuristic-based + Linear Regression	[99]
	Genetic and Search Algorithms	[100, 101]
	Graph-Based Contextual Modelling	[20]
	Hidden Markov Model	[102]
	Multi-Contextual Learning to Rank	[41]
	Naive Bayes	[45, 103]
	Probabilistic Expectation Maximization Algorithm	[104]
	Ontology-based Methodology	[105–107]
	No Algorithm	[108]

while classical relies on statistical methods [109]. Matrix factorization involves breaking down a user-item interaction matrix into latent factors that capture underlying patterns. These factors represent features shared by users and items. Classical CF, on the other hand, relies on historical interactions between users and items to find similarities. It predicts a user's preferences based on the preferences of similar users or items. Matrix factorization helps uncover hidden factors that contribute to user-item interactions. It then uses these factors to identify users or items that are alike, leading to more precise recommendations. Furthering this, Tensor Factorization (TF) takes the traditional two-dimensional MF representation and extends it into an n-dimensional problem by seamlessly integrating contextual information.

Factorization Machines (FM), introduced by Rendle in [110], are a supervised algorithm versatile in handling classification, regression, and ranking tasks. They rapidly gained prominence, becoming a popular and influential method for predictive modeling and recommendation systems. Essentially, FM represents a generalization that combines attributes of both linear regression and matrix factorization models. The advantages of FM over traditional linear regression and MF methods include: 1) The capability to model interactions involving n-way variables, where n represents the polynomial order (typically set to two) and 2) the computational efficiency for handling high-dimensional, sparse input data.

In context-aware recommendation systems, both techniques can be extended to include contextual information, resulting in recommendations that not only consider historical interactions but also adapt to specific contexts, leading to more tailored and relevant suggestions. The authors of [18] and [17] were the first to work in the idea of introducing a generalization of MF that allows for flexible and generic integration of contextual information by modeling the data as a User-Item Context N-dimensional tensor instead of the traditional 2D User-Item matrix.

In total, these techniques were exploited by 20 studies.

3.3.3 Contextual Bandits

Bandit algorithms offer a means to optimize outcomes when there is insufficient data for establishing a comprehensive statistical model. The bandit algorithm framework serves to maximize overall user engagement by striking a balance between exploiting known options (displaying the best items) and exploring novel choices (presenting new items). The exploration versus exploitation quandary asserts that exploration aids in gaining new insights about an environment or context, while exploitation involves leveraging established knowledge for decision-making or predictions [111].

This exploration-exploitation dilemma aligns closely with the requirements of a recommendation system that necessitates highlighting existing information to pique user interest, while simultaneously unearthing novel data to enhance the user experience. In many instances, these situations hinge on a user's computational patterns and the surrounding environment, making them particularly amenable to context-aware recommendation systems. This situation is perfectly illustrated in the study [59], where personalized recommendations are produced based on contextual bandits, to address the problem of deciding between trying new options and sticking with what is known. Such approach introduces a parameter-free bandit technique that uses a method called online bootstrap to continuously generate a range of estimated models. This method helps balance between exploring new possibilities and exploiting familiar ones.

In total, this family of techniques was utilized by 6 studies.

3.3.4 Neural Networks

The rise of deep learning has paved the way for innovative advancements in recommender systems. Researchers have effectively integrated deep learning models into these systems, leveraging diverse data types such as text, image, time series, etc. to enhance performance [112]. A recent focus within this field involves context-aware deep learning recommender systems, which leverage additional information to increase accuracy by considering the impact on users and items. This approach allows for a more extensive range of data to be processed, enabling the analysis of intricate relationships among these data points [76].

Although matrix factorization has proven effective for collaborative filtering, it is acknowledged that its performance may be limited by the simplicity of the interaction function, specifically the inner product. For instance, in situations where explicit feedback is employed, such as predicting ratings, it is widely recognized that improving matrix factorization models' performance can be achieved by introducing user and item bias terms into the interaction function. This modest modification to the inner product operator underscores the advantages of developing a more sophisticated and tailored interaction function to capture the subtle interplays among the latent features of users and items [113].

Following this statement, a highly cited paper aimed to adjust the inner product with a neural architecture that can learn any function from data, thereby generalizing matrix factorization under its novel framework, was [74]. Furthermore, in the work [76], the authors introduced an approach that employs neural architectures within a context-aware paradigm, leveraging numerous contextual dimensions to unveil concealed contextual patterns known as *"latent contextual embeddings"*. Consequently, a comprehensive framework for deep context-aware recommendation systems is outlined, incorporating contextual information in three distinct modes: explicit, latent, and hierarchical latent.

Since the appearance of the work "Attention is all you need" published in December 2017 [114], there has been an increasing amount of studies that investigate how to incorporate attention layers in neural network architectures, what in essence are the base of the *Transformer* model. The Transformer model has two main parts: the encoder and the decoder. The encoder takes the input data user data (like past interactions and preferences) and transforms it into a sequence of useful features like continuous representations. These are then passed to the decoder, which generates the output (recommendations) one piece at a time. Each output piece is created by looking at the whole input and the parts of the output already generated. Each layer has two parts: a self-attention layer (which helps the model focus on important data like the contextual data) and a feed-forward network (which processes the data further). There are also connections around these parts for better learning.

More specifically, an attention layer in a neural network is a mechanism that allows the network to focus on specific parts of the input data, thus giving more emphasis to relevant information. This is particularly useful in scenarios where the network adaptively adjust its attention based on the specific task, like in context-aware recommender systems. Demonstrating this point, several research papers use this technique, such as [77, 78, 80]. In particular, a recent example is found in [115], where the authors leverage context and category information to perform personalized POI recommendation, using the sparse self-attention mechanism for improving training efficiency. For instance, [116] presents a novel social recommender system based on convolutional neural networks and integrating contextual features based on text and tags. In [66], the authors consider transformers to learn bidirectional context-aware user and item embeddings with neighbourhood sampling to predict the next item that a user would interact with based on its interaction behaviour and social connections.

Among all the Transformer-based architectures like BERT, RoBERTa,T5, GPT, etc., the first one stands out in its application of CARS solutions. In this case, we find two papers that make use of BERT encoders to generate context-aware recommendations [65, 66]. In the first one, the proposed neural network is trained by jointly minimizing the loss in a layer for the labeled data, another one with consistency regularization on augmented data for user-item interaction, and finally another one with consistency regularization for user-user social context-based data. The outputs of the three layers are combined in a BERT architecture to learn bidirectional context-aware user and item embeddings with neighbourhood sampling (the samples the most influential neighbours for all the users/ items). Regarding the second proposal, the solution leverages a more complex structure beyond BERT, utilizing an LLM-based model, as we will explain next.

Further examples of the use of Transformers lead to cutting-edge solutions like Large Language models and other (complex) architectures that allows for more holistic considerations of heterogeneous data sources. These large language models (LLMs) are built on the foundational architecture of the Transformer. They are trained on large datasets, enabling them to generate coherent and fluent output. Their promising performance stems from extensive pre-training on vast corpora, followed by fine-tuning for specific tasks, allowing them to excel in various applications. When applying LLMs for recommendation, one solution presented in 2022 called "A Unified Pretrain, Personalized Prompt and Predict Paradigm (P5)" [117] stands out. P5 consolidates diverse recommendation tasks by formulating them as prompt-based natural language task into a unified framework by converting various data types—such as user-item interactions, user profiles, item details, contextual cues, and user feedback—into natural language sequences. This enables P5 to delve into deeper semantics for personalized recommendations. By employing a single language modeling objective during its pretraining phase, P5 acts as a foundational model for multiple downstream tasks. Its seamless integration with other modalities facilitates instruction-based recommendations through prompts. Leveraging adaptive personalized prompts, P5 can make predictions in a zero-shot or few-shot manner, reducing the necessity for extensive fine-tuning. Another recent example in the field of recommender system is the article [72] that makes use of fine-tuned GPT-4 to a given knowledge base and provide personalized and contextualized recommendations for cultural environments. Conversely, in the article [65], their model includes a knowledge-aware recommender system and a prompt-based conversational system. The recommender system employs context embeddings and item knowledge representation to build user profiles. They begin by encoding the dialogue history and the knowledge graph, then compute knowledge-aware user profiles, and finally retrieve items that align with the user's preferences for recommendations. The conversational system is designed to manage topics and optimize prompting vectors through contextual prompting, utilizing a pre-trained version of GPT-2.

In aggregate, 43 studies employed this methodological approach.

3.3.5 Miscellaneous

In addition to the more traditional methods described before, several advanced approaches have been employed in context-aware recommender systems. For example, optimization techniques like Genetic Algorithms (GA) have gained traction in this aspect. By considering different combinations of items and context factors, GAs can effectively adapt to changing user preferences and environmental conditions, leading to more personalized and dynamic recommendations, one example illustrating this is [118].

Other techniques to highlight are Boosting Methods, which bolster performance by iteratively gathering weaker models, making them adept at accommodating contextual fluctuations to yield more precise recommendations, as presented in [95].

Another method with particular significance in the domain of CARS involves employing Linked Open Data (LOD), ontology-based, and semantic technologies. These systems utilize conceptual representations of user needs, capitalizing on relationships within specific knowledge domains. The objective of developing such schemas is to discern the context of linked data and share common semantics, facilitating the integration of diverse content descriptions, sensor inputs, and user profiles. Notably, these technologies leverage user-generated content, such as reviews, comments, and historical activities, to enhance their functionality. In the tourism sector, they have been applied extensively to semantically interconnect data from various sources. For instance, in a study by Benouaret and Lakhoua (2015) [119], a hybrid solution is proposed, combining semantic representation of museum knowledge using ontologies with a collaborative filtering method enhanced with semantics. This system recommends artworks to users based on their past preferences, contextualized in terms of time and location. Another notable example is described in the work by Colace et al. (2020) [120], where their framework employs contextual graph approaches like ontologies, Contextual Dimension Trees, and Bayesian Networks to suggest personalized learning paths for on-site visits to cultural heritage sites such as archaeological sites.

Moreover, the combination of LOD and semantic technologies with contextual information has demonstrated to be a good solution to overcome sparsity. For example, in [39] a CB recommender where item-based contextual features are extracted from LOD, to generate an RDF graph for item modeling. Based on the same idea, in [105] a context-aware movie ontology is derived and similarity-based recommendation techniques are proposed to consider the knowledge encoded in the ontology concepts and their relationships, a method found in earlier approaches such as [97].

Moreover, Gaussian Processes offer a valuable probabilistic modeling approach, particularly advantageous for managing uncertainty, facilitating the capture of nuanced context-user preference relationships [121]. Graph representation leverages data connections, allowing graph-based methods to unveil latent patterns that translate into comprehensive and highly effective recommendations [20, 73].

Lastly, it is worth noting the utilization of Markov processes. These are models that can exist in various states with transitions between them, contingent on the current system state. While these models excel in handling sequential data and unpredictable state changes, they might not be as commonly employed as other techniques in the field. The work described in [102] illustrates perfectly this usage, nonetheless, they provide valuable insights for tackling the dynamic aspects of user preferences and item popularity. This is due to their capacity to predict the items users are likely to select next, rendering them effective for recommendation generation. These models often go by the names of "sequence-aware" or "time-aware" recommendation systems, as they depend on ordered or timestamped records of users' prior interactions as their input data.

In total, we have grouped 18 articles under this category.

3.4 Contextual techniques and methods

As motivated in the introduction, it has been shown that incorporating contextual information allows for better user characterization and profiling using more detailed data, resulting in better recommendations. Traditional approaches, such as standard collaborative filtering methods, are not suitable for these situations. In fact, incorporating contextual information, such as time and location, into the recommendation process helps alleviate the problem of lack of information. Additionally, users make decisions under certain contextual circumstances, so taking relevant contextual information into account when making recommendations helps to obtain a more accurate prediction of user preferences [5, 6].

Because of this, in this section we aim to address **RQ2.3**. Hence, to understand which contextual techniques and methods have been proposed for context-aware recommendation, we will consider the three stages of how context could be incorporated, not necessarily in multidimensional methods – which were one of the first formulations introduced [122] – but also classical recommendation approaches, through pre- and post-filtering [7]. A summary of the approaches found in this survey is presented in Figure 12, where it is evidenced that those belonging to the modeling stage are the most popular.

In the next subsections, we introduce these three stages, provide examples for each case, and discuss their main advantages and disadvantages.

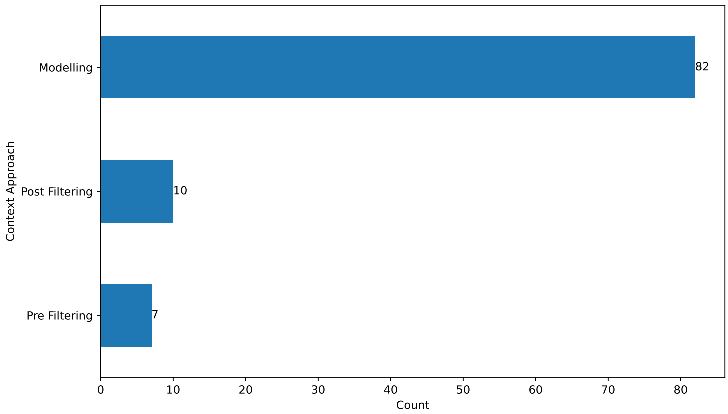

Fig. 12 Percentage of studies using each contextual technique.

3.4.1 Contextual Pre-filtering

In this paradigm, the specific context will determine the relevant information needed for subsequent canonical recommendations [13]. Essentially, a context is used for selecting or constructing the relevant set of data records, then recommendation outcomes can be predicted using any 2D recommender system on the selected subset. Therefore, only data related to context will be considered and exploited for generating the recommendation. For example, the authors from [19] proposed a novel technique called Contextual Enhanced Vector Space Model to make the algorithm context-aware by introducing a pre-filtering approach known as *micro-profiling*. To be more precise, they categorized user ratings based on the contextual situation in which the preference was expressed, and utilized this data to develop multiple context-aware (micro) profiles. These profiles were then utilized to generate recommendations that take into account the contextual situation of the user.

On the other hand, the authors of [87] proposed a Point-Of-Interest (POI) group recommendation along with multiple pre-filtering steps, with contextual factors such as rationality of the location and the intra-group influence when making group decisions are taken into place. The first step involves requirements-based pre-filtering, where POIs that do not align with the group users' preferences and therefore excludes POIs that do not comply with those requirements. During this phase, users have the option to articulate their preferences and restrictions. Afterward, a second filter is distance-based pre-filtering, which eliminates POIs falling outside the acceptable distance range for the group users. Users can specify a maximum tolerance distance in this step. Candidate POIs are then filtered based on the entire group's maximum acceptable distance. If the requirements of the entire user group cannot be met, the candidate services are filtered based on a preset distance within the system.

Pre-filtering approach was mentioned in 7 studies. Pre-filtering has often been shown to enhance diversity compared to post-filtering and contextual modeling by segmenting or profiling various hypothetical contextual scenarios, such as the item-splitting technique [123]. However, recreating the exact context scenario can sometimes be overly restrictive. Following data selection in contextual pre-filtering, data sparsity may increase, potentially resulting in insufficient data for accurate recommendations. Consequently, there remains a gap in research regarding the generalization of pre-filtering and determining the optimal number of pre-filter profiles.

These low numbers seem to indicate that its main advantage (making use of any non-contextual recommender) does not compensate its disadvantages, which are related to the lack of flexibility of such filters and its tightly connection to the domain, not allowing for general approaches applicable to diverse scenarios.

3.4.2 Contextual Post-filtering

In post-filtering approaches, contextual information is disregarded when training the algorithms, and recommendations are estimated by using a traditional 2D recommender system applied to the whole data. Subsequently, recommendations generated are adjusted via contextualization for each user by incorporating the contextual information; in other words, this mechanism refines the resultant recommendation list. The adjustments to the recommendation list can involve either filtering out recommendations deemed irrelevant in a given context or re-ranking the recommendations based on the provided contextual aspects. Similar to various recommendation techniques, post-filtering can be divided into heuristic and model-based techniques [13]. Heuristic post-filtering techniques center on recognizing shared characteristics or attributes of items for a particular user within a specific context. Conversely, model-based ones involve building predictive models to estimate the likelihood of a user choosing a specific type of item within a given context.

From the studies reviewed, we can highlight specially the work described in [33]. In that work, the authors try to overcome the difficulty of sparsity problem that happens in pre-filtering approaches by suggesting a method for the detection of the user contextual state when listening to music based on social tags of music items. The music played by the user during each session has associated tags, which are processed using word embedding techniques. These techniques are utilized for contextual post-filtering of the recommendations obtained from CF methods. Hence, post-filtering aims to find the items whose tags are closest to those of the items being played in the user's session.

In total, post-filtering has been used 10 times. Again, the advantages of this technique (which might be limited to the use of traditional non-contextual recommendation algorithms, although in this case the filters might be easier to generalize than for the pre-filtering case) are not strong enough to overcome its disadvantages (related to the additional effort required by the algorithms to discriminate among all the information received, since contextual information was removed). While some studies advocate for the implementation of post-filtering to reorder recommendation lists, thereby ensuring that recommended information aligns better with the current scenario, this approach incurs a computational cost.

3.4.3 Contextual Modeling

Contextual preference estimation and elicitation methods have garnered significant popularity and attention within the academic community from the outset, aiming to integrate context seamlessly into existing recommendation solutions. Diverging from earlier methods, this approach directly integrates contextual information into the recommendation function [13]. This can be achieved through predictive models or heuristic calculations that consider contextual information alongside user and item data. Consequently, this necessitates a shift from the traditional user-item-rating data input to a user-item-rating-context input, resulting in multidimensional recommendation functions that have the potential to offer more accurate recommendations.

Among the three categories, this approach is the most prevalent, with almost all the reviewed studies applying this context-aware approach. For instance, one of the pioneering and notable papers that addressed the context-aware challenge is the work by Karatzoglou et al. [17]. That paper leveraged the emergence of MF techniques, which gained popularity through the Netflix Prize competition [124]. Based on their methodology, the Users $\times$ Items $\times$ Contexts space is modeled as an n-dimensional tensor, and the factorization of this tensor provides a concise model for generating context-aware recommendations.

As mentioned previously, a significant emphasis in the field of CARS is placed on contextual modeling. This is largely attributed to the advancements in neural network architecture, which have facilitated a more refined and seamless integration of context. In total, 82 studies have proposed or exploited contextual modeling. This number highlights that, despite its primary drawback—requiring significant domain knowledge for appropriate modeling—these techniques yield promising results. Their ability to adapt to specific scenario characteristics, typically by integrating context and intuitively capturing interactions among users, items, and context, underscores their effectiveness. Contextual modeling, renowned for its capacity to enhance recommendation performance, emerges as the leading paradigm for integrating context into recommendation systems, underscoring its advantages in modeling and leveraging contextual information.

3.5 Application domains

In order to address the research question **RQ2.4**, an analysis was conducted on the application domains of context-aware recommender systems as presented in the reviewed papers. The findings regarding the categories of these domains are summarized in Figure 13. The categorization criteria for different domains were determined by the researchers based on the descriptions of the respective domains and datasets.

For example, the criteria for the multimedia category included elements such as movies, songs, actors, direction, and genre tags. In the case of e-commerce/ads, the criteria encompassed product type, brand, size, and time of purchase, among others.

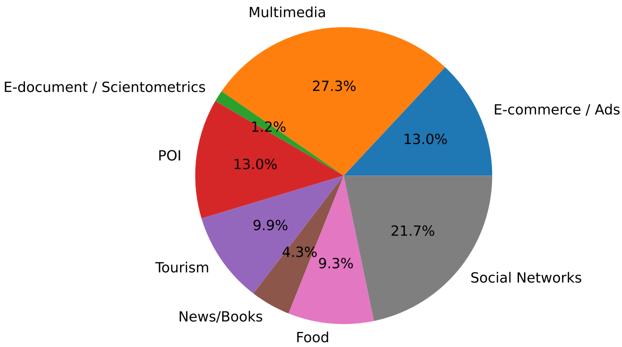

Fig. 13 Percentage of studies whose experiments belong to each of these domains (A study could be multi-domain).

The analysis reveals several notable observations. Firstly, it is evident that the domains of Multimedia, Social Networks, and E-commerce/Ads feature prominently in the reviewed papers, indicating their dominance over other categories. Approximately 71 out of the 90 studies. fall within these two categories. Following closely behind is another cluster comprising POI (Points of Interest) and Tourism [3], which are discussed in 21 and 16 studies respectively. On the other hand, the domains Food, E-document/Scientometrics and News/Books are less popular, with only 22 identified studies in total.

Finally, it is noteworthy that the majority of the examined studies concentrated on a single specific domain, as illustrated in Figure 14, with 48 out of the total 90 papers. There is an observable declining trend, wherein as the number of domains increases, the number of studies involving multiple domains decreases.

3.6 Evaluation baselines

In the realm of recommendation systems, baseline models hold a crucial position, serving as benchmarks against which the performance of algorithms can be evaluated. They not only quantify the extent of improvement but also facilitate comparisons with previous methodologies. This performance difference is crucial in assessing the efficacy of proposed, more sophisticated methods. Additionally, the selection of an appropriate baseline must consider factors such as implementation effort and system maintenance requirements. Thus, a baseline model must meet certain criteria to encompass these considerations.

By selecting accurate baselines, the advancement of recommender systems can be facilitated. However, in recent years, the proliferation of models for experimentation has posed challenges. Researchers often limit the number of baselines due to factors such as computational complexity, particularly in the realm of deep learning approaches [126]. This limitation can result in a trade-off between including

[3] Following [125], we analyzed separately places (POIs) and tourism domains.

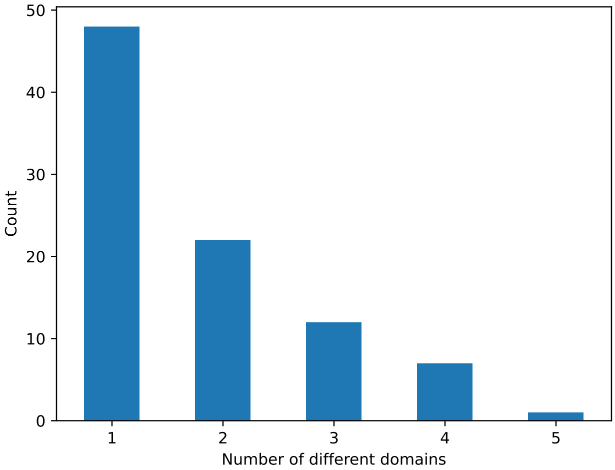

Fig. 14 Number of domains applied per study.

comprehensive algorithms and the constraints of time and resources available for experimentation.

As highlighted in [127], an ideal baseline should possess several desirable characteristics: minimal implementation effort, competitive performance, and the absence of additional data requirements. Baselines that consistently fail, are difficult to implement, or necessitate additional data are impractical for experimental use.

Furthermore, the choice of baselines should consider the domain in which they are applied. In the domain of CARS, baselines can be classified into traditional non-contextual versus context-aware approaches. Context-aware baseline models are particularly pertinent in studies where contextual information is deemed valuable, aiming to showcase the benefits of more sophisticated techniques. To address **RQ3.1**, we conducted an analysis of the studies within this review and categorized them based on the inclusion of non-contextual baseline models in the experimental comparisons. The majority of studies (approximately 71.1%, as depicted in Figure 15), utilize non-contextual baseline models. Among these traditional model baselines, the most frequently employed ones include Random [59], Popularity [70], Matrix Factorization [21], and Bayesian Personalized Ranking [25].

However, there are instances where traditional non-contextual baselines are not employed. For instance, in [35], Tensor Factorization serves as an independent contextual model baseline, compared against models utilizing context similarity as an alternative contextual modeling approach, and integrated into sparse linear methods and matrix factorization algorithms. Similarly, in [61], an auto-encoder is trained to predict user preferences based on past interaction contextual data, and evaluated against machine-learning-based models such as Random Forest or Decision Trees.

In general, the majority of analyzed papers claim the superior performance of their proposed solutions over baselines, regardless of contextual or non-contextual conditions. However, there are instances where the desired level of accuracy performance is not achieved. For example, in the study [106], a novel approach to finding music recommendations based on users' context via semantic relations was proposed. Although their system exhibited improved serendipity, it was not successful in simultaneously achieving improvements in accuracy and serendipity.

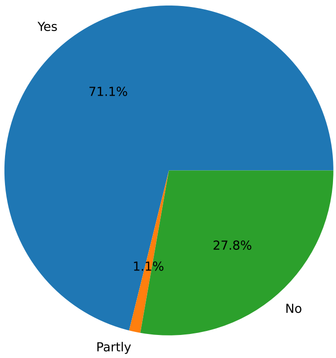

Fig. 15 Percentage of studies where a non contextual-aware baseline was used.

3.7 Evaluation metrics

The aim of RS evaluation is to determine which recommenders are better than others based on the results obtained in certain metrics under a specific evaluation methodology. In fact, among the different types of experiments that can be performed with users of a particular RS – that is, offline, online, and user studies [128] –, the RS community has been mostly focused on offline evaluation, since it is the most comparable one across different settings and the one typically used in the literature. This also occurs in CARS, as evidenced in Figure 16 as a way to address **RQ3.2**.

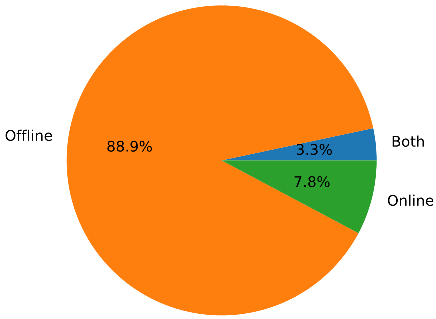

Fig. 16 Percentage of studies showing each type of evaluation used.

For this classification, we follow the same distinction made in [128]: an offline experiment is performed by using a pre-collected data set of users choosing or rating items, using this data set we can try to simulate the behavior of users that interact with a recommendation system; a user study is conducted by recruiting a set of test subjects, and asking them to perform several tasks requiring an interaction with the recommender system, while the subjects perform the tasks, we observe and record their behavior; whereas an online evaluation is considered when the system is used by real users that perform real tasks, where typically such systems redirect a small percentage of the traffic to different alternative recommendation engines, and record the users interactions with the different systems.

In our analysis, we found no examples of works considering user studies, and less than 8% of the works used exclusively online evaluation. One work where an A/B test was used inside of Amazon Music was [57]. Given the small number of this type of evaluation, it should be noted that we also considered offline evaluations where a realistic simulation based on historical logs was used as online, as in [59]. Classical offline evaluation are easier to find, like the works [17, 20, 40, 82, 92]. It is interesting, however, to note the works where both evaluations were reported: [129] and [55]. In the first case, it is not a real online experiment, as the authors report *online recommendation efficiency* to measure the time cost of the recommendation approaches if they were deployed. The second case is similar, where the online part of the evaluation refers to the way the algorithm learn (in an online fashion). Thus, we conclude that none of the works included in our study *actually* consider more than one type of evaluation.

Depending on how the evaluation is devised, we need different ways to analyze the performance of the recommenders. Originally, the recommendation quality was equated to how close the recommender was able to predict the rating provided by the user. Hence, error metrics like Mean Absolute Error (*MAE*) and Root Mean Squared Error (*RMSE*) were used, however, since these metrics only account for the observed items, they do not reflect well real-world problems nor the perceived user experience [130]. Because of this, Information Retrieval ranking metrics like *Precision*, *Recall*, or *nDCG* (normalized discounted cumulative gain) were used to measure how many relevant items were included in the ranking generated by the RS, usually, up to a ranking position or cutoff [131].

As summarized in Figure 17, ranking metrics are the most frequent ones, although error metrics are still reported. Examples of works where only error metrics are used include [17] or [46]. These works tend to be older (as the community as a whole is replacing these measurements with more appropriate ones), in fact, the most recent works using *MAE* or *RMSE* include [100] and [84], evidencing this paradigm shift. Nonetheless, some recent approaches can also be found, although they no more report only error metrics, but in combination with ranking metrics, like [121] and [78]. At the same time, those works running an online evaluation used other types of metrics, such as *CTR* (Clickthrough rate) or lift percentage in key performance metrics (like successful playbacks or users with successful playbacks), see [57, 59].

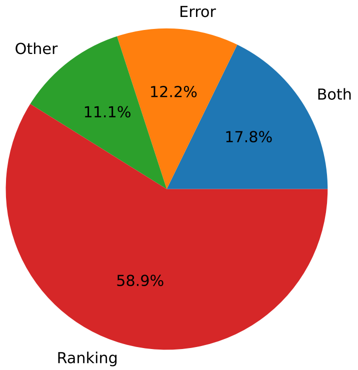

Fig. 17 Percentage of studies where each type of evaluation metric was used.

More specifically, Figure 18 shows the frequency of each evaluation metric. It further emphasizes the prominence of ranking metrics such as *nDCG*, *Precision*, and *Recall*. We also observe the long tail of less popular ranking metrics like *AUC*, *average rank*, or *ERR*. It is important to highlight that most of these metrics need a cutoff value to be computed, and this was not considered when counting how often they were used, otherwise the overlap would be much lower since each paper may use very different cutoffs. Additionally, except one case (*popularity rank*) all the metrics are tailored for accuracy, neglecting other dimensions like novelty or diversity [132].

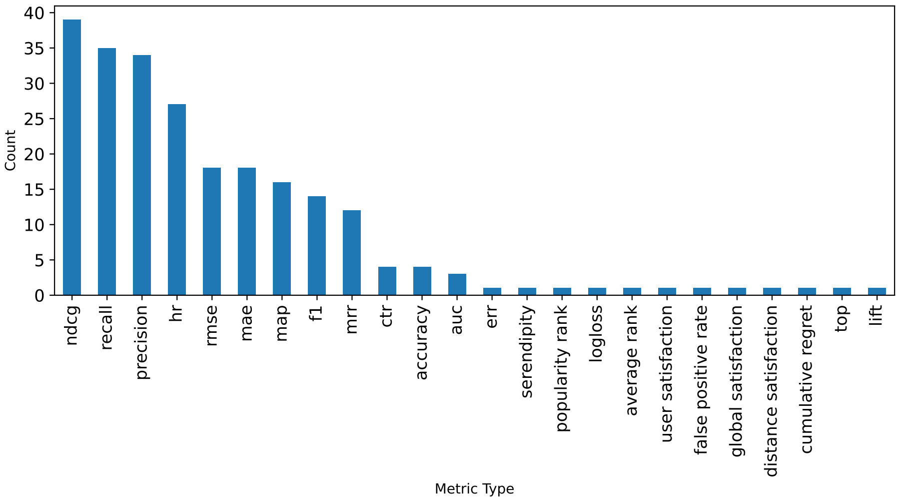

Fig. 18 Number of studies where each evaluation metric was used.

3.8 Evaluation datasets

In line with **RQ3.3**, we analyzed the datasets exploited for conducting the experimental evaluation of the techniques described in the reviewed studies. We summarize the studies for each dataset that was considered in more than one work in Figure 19, merging and summing all the other occurrences under the 'other' bar, meaning that more than 70 datasets were considered only once, such as Citeseer [89], YouTube [90], or KDDCup12 [59].

The most popular datasets (Movielens, Lastfm, and Yelp) belong to three different domains (movies, music, and food/social media) and provide simple, but standard contextual variables: time and item tags, and for the case of Yelp, also item location. The Frappe dataset belongs to the mobile app recommendation domain and provides temporal and location contextual information. The food dataset is composed of ratings on menu items and includes two contextual factors: one describing if the situation in which the user rates is virtual or real, and another that captures how hungry the user is. The CoMoDa data is a publicly available context-aware movie data collected from surveys; there are 12 context dimensions that captured users' various situations, including mood, weather, time, location, companion, etc.

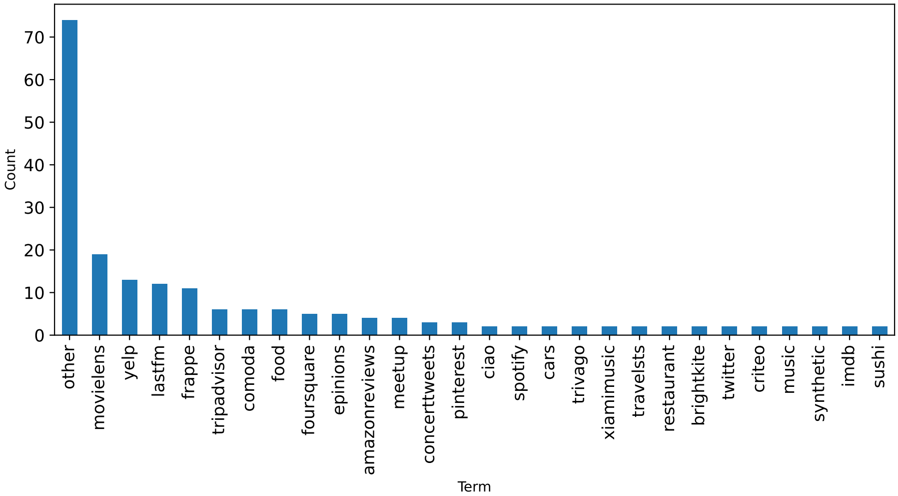

Fig. 19 Number of studies where each dataset was used.

TripAdvisor and Brightkite belong to the POI domain and include temporal and geographical contextual features. Epinions is a review dataset that includes who-trust-whom online social network, its main contextual feature is user trust. Amazon reviews contains reviews from the e-commerce domain and its main contextual attribute is time. The Meetup dataset is a sample of the well-known event-based social network and allows capturing social, spatial, and temporal contextual features.

Foursquare provides user interactions with POIs, including temporal and geographical information. Pinterest includes implicit feedback data and, like in Amazon reviews dataset, its main context is time. Among the least used datasets, we focus on ConcertTweets and Trivago. ConcertTweets is created based on publicly available and well-structured tweets referring to music concerts, it contains geolocation information about the concert and timestamp information about the item and the rating event. Finally, the Trivago dataset was released in the context of the 2019 RecSys Challenge to recommend hotels in booking sessions, it included ratings, queries, and other interactions belonging to the same user session, which could be considered as contextual features, together with the country platform, used device, and prices of impressed items.

3.9 Reproducibility considerations

In relation with **RQ3.4** and QQ8, we summarize in Figure 20 how many open source implementations exist in the analyzed studies, evidencing that a majority of the papers do not provide an implementation.

It should be noted that for this statistic, we only considered the cases where the paper reported an implementation, not when a public implementation of the proposed method was made available afterward (often, from other authors). For example, [78] does not mention any repository in the published paper, but the repository https://github.com/alexmay21/AIN claims to include the implementation of this paper and at least one of the authors matches the username of its commits.

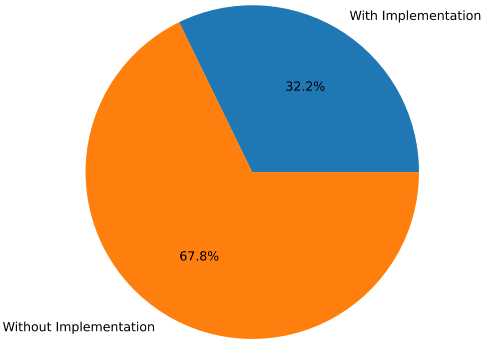

Fig. 20 Percentage of studies where an implementation was provided.

Besides that, as Figure 21, shows the number of studies that incorporates an open-source implementation has increased substantially throughout the years.

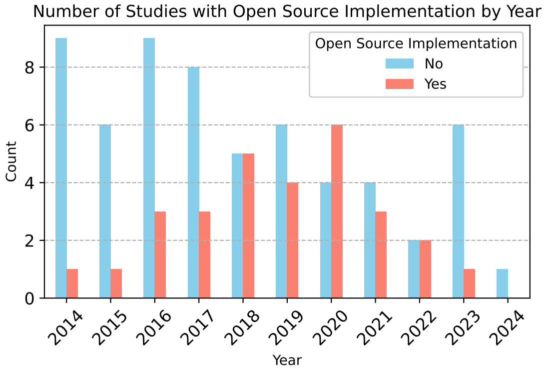

Fig. 21 Number of studies with/without open-source implementation per year.

3.10 Future works

In the following, we analyze the suggestions and discussions for future works mentioned in the reviewed studies in order to provide an answer to **RQ4**.

Several authors emphasize the idea of *enhancing the contextual space*. This future work is related closely to the problem of understanding context relevance and recommendation quality. Some suggestions include development of dynamic models and exploiting more the real-time recommendation and predicting the user intent at the moment like [75].

A recurrent future line of work is the necessity of *understanding the user behaviors, preferences, and intents*. This is fundamental for the goal of a better personalization and user modeling. Some suggestions involve incorporating the user in the recommendation process by informing their intentions [95], building user-centered metrics to strengthen the model performance as in [19], or incorporating social context information, which implies understanding how users are influenced by their social networks and tailoring recommendations accordingly [49].

Another reported potential possibility is to improve the contextual recommendation by using *diverse data sources and context information* to expand the recommendation capabilities as well as incorporating *multi-criteria ratings* for a more comprehensive understanding of user preferences. For example, by investigating multi-modal recommendations, i.e., incorporating factors like images and user behavior beyond text data [32].

As a typical issue in CARS is their limited scalability, *model optimization and scalability* is a problem that has been stood out in different studies. Some works mention as potential ideas to address this problem to optimize models in the prediction phase by researching the adoption of mini batch learning (where the model makes adjustments based on a smaller group of training examples). This includes the most extreme case, known as online learning, where the settings of the model could be updated after every single step and therefore analyze how the models perform in online dynamic settings with burst streaming workload [85].

Almost every future work section considers the *extension or modification of the presented solution*. Due to the high rise of deep learning techniques applications, including the use of Large Language Models, the most common suggestion is to research about the use of new and more optimized architectures (such as transformers and attention layers), plus digging into the use of dynamic graphs based on contextual information [29]. Conversely, exploring different loss functions to enhance recommendations, improving both ranking and recommendation, or exploring ensemble methods that combine models optimized for various aspects of recommendations present themselves as potential solutions, see [21, 63].

Other works aim to *extend the contextual recommendation models to different domains* and applications [79]. In particular, this would imply to evaluate the impact of these models on business performance metrics, beyond traditional evaluation metrics, or integrate explainability approaches into CARS [64].

4 Discussion

In the following, we discuss the outcome of our systematic literature review, considering the answers provided in Section 3 to the research questions originally introduced in Section 2.1.

4.1 Included studies

As reported in Figure 3, the earliest studies included in this literature review are from 2014. Indeed, until the year 2015 only conference publications matched the criteria considered in our analysis. During the range 2015-2019, a large portion of the studies included in our review were published, still mostly led by conference venues (like ACM RecSys and ACM SIGIR) but with more journal papers being published.

Regarding the quality of the included studies, summarized in Figure 4, we can highlight the fact that higher scores were assigned to works published in journals and conferences with respect to workshops, while the quality of journal and conferences were comparable to each other.

4.2 Types of context

The taxonomy presented in Section 3.2 reveals that the concept of context is inherently ambiguous. Depending on the specific focus, the notion of context can exhibit various interpretations. The initial aspect discussed pertains to acquisition (Section 3.2.1), where it is evident that both categories (interactional or representational) coexist in nearly all the examined studies. This aspect is closely interlaced with the comprehensive goal of these papers, which is to assess algorithms across diverse domains, entailing the involvement of distinct data sources. Additionally, it is essential to acknowledge that the traditional approach to CARS exclusively considers a representational perspective, encompassing predefined and fixed contextual factors like time, location, or companionship.

However, this representation does not always align with reality, particularly when dealing with factors that are not easily accessible or observed, such as a user's mood or intent. Furthermore, in certain scenarios, context only becomes apparent upon the completion of an action, neither preceding it nor existing independently of it.

In contrast, the interactional perspective defines context as a dynamic relational property that emerges between activities. This view rejects the notion of enumerating relevant contextual factors prior to a user's activity initiation. This poses a significant limitation in practical applications, where the representational view falls short. Over time, researchers have increasingly recognized the disparity between traditional context modeling and real-world scenarios. Consequently, they have begun incorporating the interactional dimension of context, assuming that user behavior is influenced by an underlying but unobservable context. This kind of approaches enables the development of solutions that harness the strengths of both representational and interactional views.

An interconnected yet distinct issue pertains to the concept of observability (Section 3.2.2), specifically regarding the depth of knowledge concerning contextual factors, their structure, and values. This issue arises when contextual knowledge exists in a phase separated from the target domain and must be transferred to benefit the target domain where recommendations are made. This challenge can manifest in situations where context is initially acquired *representationally* and is fully discernible at the outset, but is subsequently transformed into a latent, concealed factor. For instance, in the work [33], music tags associated with the music played by the user in each session are subjected to word embedding techniques to facilitate contextual post-filtering of recommendations generated by CF methods. This approach seeks to address the issues raised by the initial classification framework and address the limitations previously encountered, hence relying on the categorizations of fully observable, partially observable, and unobservable contexts. Conversely, context is considered unobservable when relevant information is entirely unknown or inaccessible, as is the case with users' intent. Consequently, it cannot be explicitly utilized for recommendation purposes. Nonetheless, the primary concern in these instances is that unobservable context gives rise to temporal dynamics in user-generated data, needing consideration in order to provide more precise recommendations.

Lastly, context can be assessed through its dynamism (Section 3.2.3), considering whether and how its availability and structure evolve over time. The majority of the selected studies incorporated dynamic contextual factors in conjunction with static ones or independently. This holds particular significance since real-world scenarios are consistently exposed to dynamic and evolving factors, thus exerting an ongoing influence on users and items. Omitting this aspect from the equation renders the approaches less adaptable and less suited to real-world situations, as they fail to encompass the complete context.

4.3 Recommendation approaches

Around half of context-aware recommender systems fall into the category of Neural Network approaches, as depicted in Figure 10. This prevalence is primarily a consequence of the proliferation of deep learning architectures and their appeal to the academic community, primarily for their capacity to enhance recommendation task performance.

Upon delving deeper into this subject, we can emphasize the keen interest in integrating attention layers within deep learning settings. Within these attention networks, it is possible to input embeddings from cutting-edge deep learning models specifically tailored to each data type. Consequently, the attention networks can adapt their focus based on the intricate contextual attributes within the user's preference hidden representation. This acknowledgment is evidently demonstrated in Table 2, where the trend is clearly visible.

It is important to note that these approaches often do not solely rely on neural network techniques but rather incorporate a fusion of classical methods. Therefore, the majority of recommender systems can be considered as hybrid systems, blending multiple techniques for optimal performance. A remarkable illustration in this context involves the amalgamation of MF and TF with various Neural Network architectures across different combinations, also shown in Table 2.

Last but not least, the other half of the research focuses on recommender systems that can be classified as classical collaborative filtering, factorization machines, or miscellaneous techniques. This choice arises from the design of systems rooted in domain-specific applications, where simplicity and domain-specific insights lead to improved results.

4.4 Contextual techniques and methods

In Section 3.4, it becomes evident that the majority of context-aware approaches fall under the category of contextual modeling. In this sense, contextual information is seamlessly integrated into the recommendation process, offering several advantages in addressing challenges such as more accurate estimation, enhanced context extraction, and improved variable identification, and explaining why these approaches may advance the research field, in contrast to the rest of techniques and methods. Furthermore, as demonstrated in the study conducted by Panniello [133], contextual modeling strikes a suitable balance between accuracy and diversity when compared to its counterparts, such as post-filtering and pre-filtering techniques.

With the rise of neural network-based techniques in recommendation systems, researchers have shifted their focus toward enhancing the recommendation process using these architectures. In many cases, these approaches have proven to be more effective than traditional methods and both pre-filtering and post-filtering techniques. However, as the diversity of context types to consider increases, there is a pressing need for a recommender system that can efficiently incorporate contextual information while mitigating issues related to data sparsity. When similar users or items share similar contextual information, there is an opportunity to enhance the performance of existing recommender systems by reflecting these contextual features.

For instance, in the research conducted by Unger et al. [76], a deep learning framework was proposed to define context as any user situation that may influence user preferences. By using these contextual dimensions in such algorithms, it is possible to uncover latent contextual patterns, referred to as latent contextual embeddings, without the need for predefining them. This study demonstrated that leveraging structured latent contexts within the deep recommendation framework yields significantly improved performance compared to other context-aware models across all tasks. Hence, it is their flexibility in combination with their good results that make modeling approaches the most appropriate candidates in most scenarios.

Nevertheless, the literature reviewed here shows a limited number of studies exploring pre-filtering and post-filtering systems. The pre-filtering approach, which was extensively used in the past due to its association with static factors and its ability to reduce computational complexity by independently segmenting user-item ratings based on contextual scenarios, has gradually lost prominence. Besides the notable interest in contextual modeling research, pre-filtering entails additional efforts in acquiring information, as it requires predefined contextual factors and user-item matrices.

Furthermore, pre-filtering has primarily been employed alongside the Representational-Static approach for modeling contextual factors, expanding its application to more intricate context modeling scenarios, including latent and dynamic cases. However, this emerging line of research remains largely unexplored, with only a handful of studies investigating the integration of dynamic features, as demonstrated by [54, 62].

Similarly, the post-filtering approach has seen a decline in research attention. Although it seems that this technique may hold promise for future investigations, where contextual information is not only used to filter out irrelevant recommendations in a given context but also to re-rank recommendations based on the given context, it has received less focus. For example, the work by Aliannejadi and Crestani [104] demonstrated that predicting contextually suitable locations and re-ranking suggestions based on their contextual appropriateness leads to more accurate top-k location recommendations by hybridizing the post-filtering step with contextual modelling.

Finally, it is worth noting that some studies amalgamate the best of both worlds by combining various strategies, such as contextual modeling with pre-filtering and/or post-filtering. To exemplify this, the paper authored by Jawarneh et al. [75] introduces a hybrid algorithm that adapts and repurposes a pre-filtering contextual integration method, integrating the resulting new dimension into a deep learning-based neural collaborative filtering approach. This approach retains and capitalizes on the strengths of both methods while mitigating their respective limitations.

4.5 Application domains

As demonstrated in Section 3.5, the most commonly targeted domains include multimedia, e-commerce, and a diverse category encompassing items such as food, social media, and news. This outcome is quite definitive and aligns with the issue previously discussed by Adomavicius and Tuzhilin [134], which pointed out the limited applications of CARS in this specific field due to the scarcity of publicly available CARS datasets. While major companies like Amazon, Spotify, and X (formerly known as Twitter) have undertaken initiatives in recommendation services, they often release minimal information about the datasets employed, hampering the research community's ability to replicate these experiments accurately.

The majority of the recommender systems examined in the literature are domain-agnostic, as they rely solely on publicly accessible datasets. However, some solutions are tailored to specific domains due to data collection methods, such as web scraping or querying specialized API services like Meetup or Medium. Consequently, the reproducibility of these domain-specific systems becomes challenging. As a result, the choice of application domain is heavily contingent on the sources of these datasets. For instance, the widely recognized MovieLens dataset places *multimedia* in the spotlight among various categories.

It is worth noting that the *movies* domain is unique in that it encompasses a broad range of context types, including time, social interactions, user activity, and location. This presents researchers with a significant opportunity to evaluate multiple contexts simultaneously without the need to generate domain-specific datasets. This situation is analogous to other domains like *music* and *e-commerce*, where context elements like user activity, location, and time are also present. An example of this is evident in studies utilizing LastFM as their dataset for testing.

While location is a contextual attribute that, as shown, can be exploited in many domains, it achieves critical importance in the POI and tourism domains, in particular, under the so-called Location-aware recommender systems [30]. These cases may consider other information sources such as social connections, weather, and sequentiality, but location and other inherent properties of the items in the city, allow for a broad range of techniques to be applied [31, 135, 136].

Conversely, domains such as *books*, *e-learning*, and *e-documentation* are less explored in the context-aware recommender systems landscape. This is likely because establishing detailed context scenarios for evaluation purposes is a more complex task, making it challenging to locate well-curated data for testing. Additionally, these domains represent niche categories within the real-world application spectrum.

4.6 Evaluation baselines

In Section 3.6, we observed that non-contextual models are predominantly utilized in studies to demonstrate performance improvements. However, adherence to this evaluation protocol varies due to the absence of rigid guidelines outlining a comprehensive set of essential baselines. This variability in baseline selection in the field of CARS reflects the diverse objectives pursued by practitioners.

While accuracy performance evaluation often relies on established baseline algorithms, assessing algorithm robustness and beyond-accuracy metrics presents greater ambiguity and is less common in terms of non-contextual baselines. For example, in the study by [49], the authors noted a scarcity of research exclusively focused on the robustness of context-aware recommender systems. To address this gap, they introduced adversarial perturbations during training to enhance model robustness, despite not presenting comparisons against traditional baselines.

Another consequence of deviating from non-contextual baseline models comparison is evident in subsequent research conducted by the same authors. For example, in [35], the authors compared their proposed methods against contextual modeling approaches, having already established the effectiveness of previous model versions in earlier works such as [42].

It is noteworthy that among non-contextual baseline models, a significant disparity exists when considering deep learning approaches. As discussed in [126], there is concern regarding the performance of complex neural methods, with 11 out of 12 reproducible approaches being outperformed by conceptually simpler methods. This phenomenon is attributed to the inadequate choice of baselines, leading to a distorted perception of consistent results, and the tendency for papers to be rejected if they do not demonstrate improved accuracy performance. This trend is also observed in CARS, where only a few traditional methods are selected to demonstrate the benefits of new state-of-the-art models, with comparisons primarily focused on neural-based approaches. As recommended by authors of the latter one, including baseline algorithms from various families during the experimentation stage is essential. Additionally, systematic optimization of algorithm benchmarks, encompassing all relevant parameters and baselines across datasets and metrics, is crucial for meaningful comparisons and assessments.

Finally, it is important to recognize that while the inherent aspect of personalization within the concept of contextualization is assumed (e.g., relating specific contextual scenarios to past user activity will align more closely with user traits), the comparison between non-contextual and contextual methods relies solely on abstract metrics. As demonstrated in the study by Braunhofer et al. [137], a live user study was conducted to evaluate how the relationship between "personalization" and "context" enhances baseline performance. This suggests that new research endeavours should prioritize (when feasible) providing stronger evidence regarding personalization, such as through online experiments or live user studies.

4.7 Evaluation metrics

As reported in [128] and summarized previously in Section 3.7, the types of experiments that are usually performed within recommender systems are classified into offline, online, and user studies. CARS were often evaluated using offline comparisons, as seen in Figure 16. This result is expected, as such an approach is less expensive than performing a user study or running online experiments. Nonetheless, as we shall discuss later in Section 4.9, recommender systems evaluation is a complex task, and there are several stages in the process where decisions might have a strong impact in the final result, even when using standard datasets and well-known evaluation methodologies [138]. Hence, the strong focus on offline methodologies might be one solution towards comparable results, but, as we shall discuss later, this is not enough. At the same time, it is worrying that no user studies were found in our literature review, as it prevents from understanding the real impact context-aware recommendations may have on users.

In terms of evaluation metrics, the majority of the reviewed studies validated their proposals using ranking metrics, as presented in Figures 17 and 18. This result is probably linked to the shift in the community from error-based metrics to ranking ones, acknowledging the importance that presentation has in the overall user experience. Even though our analysis included a category called 'other', we want to emphasize that no metric measured concepts unrelated to accuracy, hence neglecting all the recent literature on novelty, diversity, and fairness that is pervasive in the area [132, 139]. One possible reason for this scarcity could be the lack of standard datasets (see Section 4.8) and evaluation methodologies applicable to a majority of CARS.

4.8 Evaluation datasets

From the results presented in Section 3.8, we observe that there is no standard dataset in context-aware recommendation, since almost each work uses its own dataset. This is in line with previous observations from the literature [140]. This issue is related to the multifaceted aspect of this research area: as discussed in Section 4.5, context can be exploited in almost any domain, opening up several possibilities in terms of collecting or representing the context as part of a dataset. Except for a few datasets that already existed in the literature, in many cases the dataset had to be built for the study, reinforcing this issue of no standard collection. This, to some extent, might be reasonable, as the nature and definition of context is too diverse (see Section 3.2), which would make it virtually impossible creating a dataset that could be applied to any type of research endeavor.

To dig a little bit deeper on the datasets used by the community, we have collected all the datasets reported in the analyzed studies, to summarize their main characteristics. First, when doing so, we realized the number of contextual dimensions is often not explicitly mentioned in the papers (sometimes even the number of interactions or users is not reported), but it has to be inferred based on other information. In fact, in some papers, it is not easy to discern if the context is used in prediction time or when modeling the user or item. In particular, when context is observable, the number of dimensions tend to be lower than 5, whereas in other cases it may reach higher dimensions, as it is not really limited by the data but by the latent variables being modeled, as in [55], where context is built from a TF-IDF vector after a PCA process. Another common situation where the number of contextual dimensions is unknown occurs when context is created from the last items interacted by the user, as in [71] and [88].

Second, large datasets are very difficult to obtain, and authors tend to complement already publicly available datasets with external information to include contextual attributes. For example, in [46] the authors extend a dataset based on Epinions by creating two contextual features: age and trust, although the most typical dimension being computed is time, as in [53], where the temporal dimension is divided into a predefined number of bins and each rating event is assigned to the respective bin, or in [23], where the time and order of transactions were derived from data.

Third, it should be emphasized that in a very small number of cases, the context variables were artificially generated. For example, in [100] the demographic context was randomly generated, whereas in [17] the authors replaced the gender feature with a new artificial feature. Nonetheless, exploiting synthetic generators for the evaluation of CARS seems a promising approach to address the lack of datasets; in fact, some authors used such tools to complement experiments based on both real and synthetic data [54, 141].

4.9 Reproducibility

As shown in Section 3.9, most of the studied works did not provide any implementation. This, per se, should not be a burden for reproducibility, as long as enough details are provided in the research paper, such as the data and any processing steps followed, the baselines and their parameters, the evaluation methodology, and so on [138]. Moreover, as discussed in the previous section, no standard datasets exist (yet) for CARS, so this makes reproducing previous research more tedious and difficult.

Table 3 Datasets with the same name but different statistics (taken from the corresponding article).

Dataset	Used in	Users	Items	Interactions	Contexts
Foursquare	[77]	10,766	10,695	1,336,278	1
	[129]	20,293	123,968	334,287	2
	[100]	183	304	518	3
LastFM	[23]	992	174,091	18,908,597	2
	[33]	53	86K	420K	last n
	[40]	983	60,000	246,853	1
	[49]	2,917	1,853	219,702	1
	[85, 142]	1,000	20,301	214,574	7
	[55, 80]	1,892	17,632	92,834	1
MovieLens	[52]	138,493	19,011	19,854,570	3
	[25, 46, 68, 85, 92, 121, 142]	6,040	3,952	1M	5
	[79]	671	9,066	100,004	1
	[46, 80, 92]	943	1,682	100K	5
	[49]	693	2,634	30,830	1
Yelp	[78]	96,143	49,482	2,283,913	4
	[77]	38,945	34,245	981,379	1
	[25]	25,815	25,677	730,791	1
	[80]	16,239	14,284	198,397	1
	[29]	6,336	13,003	185K	5
	[29]	5,170	12,997	144K	5

In fact, we have collected in Table 3 a sample of the datasets that appeared more frequently in our analysis under the same name but, as it is obvious from the table, where they do not correspond to the same data because their main statistics (number of users, items, interactions, or contextual dimensions) do not match as reported by the authors.

On top of that, in contrast to classical recommender systems, there are not so many libraries or frameworks focused on context. This forces every researcher to start their implementations from scratch, taking decisions and introducing potential deviations from other studies.

Finally, establishing reproducibility should be considered a standard in applied machine learning. Over the years, there has been an increasing trend, likely due to reproducibility becoming an evaluation criterion for conferences, as depicted in Figure 21. This offers an advantage by enabling other researchers to reproduce results to a certain extent. However, the community should strive to elevate the level of reproducibility by incorporating hyper-optimization, evaluation, data processing, and baseline models. Otherwise, validating reported results becomes more challenging.

4.10 Future works

In Section 3.10, we outlined the primary future directions emphasized by the authors of the reviewed studies, as well as the key aspects regarding the progression of the field of context-aware recommender systems.

The most commonly suggested approach to enhance current techniques involves expanding the contextual space and utilizing various data sources. This suggestion aims to improve personalization, enhance user comprehension, and identify user intentions more accurately. However, it would require more data, one of the main problems in this area, as highlighted in previous sections.

It is essential to note that one of the significant challenges highlighted by the authors is the scalability of the existing solutions. This challenge conflicts with the idea of expanding the contextual information space, since having more information available usually requires more resources.

Additionally, a recurring solution proposed in many papers is the idea of customizing the presented solutions. This highlights the importance of focusing on designing better learning algorithms for specific tasks, rather than relying solely on intricate and costly methods for marginal improvements. As a particular approach towards customization, several authors express their intention to dedicate more effort to extending their solutions to different domains, which comes as a result of CARS are becoming more task-specific compared to their initial purpose. Considering the architectures and Language Models experimented so far (including transformer, convolutional or recurrent neural networks, and LMMs based on small models), novel techniques such as fine-tuning (especially for textual content) or arquitectures and optimization approaches (graph neural networks or two-tower architectures) should be considered to better adapt to specific contextual tasks.

Furthermore, the evaluation of recommender systems is a complex undertaking, as discussed earlier. Therefore, various authors express their plans to conduct more extensive experiments on the proposed approach in future works. This occurrence is often due to the challenges in establishing standardized experimentation frameworks, primarily attributed to the specific domains under study. We argue this is also a consequence of the lack of standard datasets and public implementations of well-known approaches, as evidenced in our reproducibility analysis.

5 Conclusions

Summary of main observations and lessons learned

In this work, we conducted a thorough examination through a methodical review of 90 studies spanning from 2014 to 2024. By doing this, we aimed to contribute insights into context-aware recommender systems. The primary objective was to aid the research community in comprehending the effective integration of context into the recommendation process. Through the four research questions formulated herein, our objective was to uncover the real issues in the recommender systems field, the types of exploited context, the commonly employed recommendation approaches, and the use of evaluation methodologies across different domains.

A noteworthy observation relates to the insufficient coverage in understanding the concept of context and its modeling in recommender systems. The multifaceted nature of context, often perceived as an ill-defined and complex issue, stems from ongoing debates surrounding its conceptualization. While some frameworks are inching towards a common point, there remains a need for a unified set with a standardized definition that can be universally embraced.

We categorized the recommendation approaches mentioned in the surveyed works based on the algorithm families they fall into. Our observations indicate a predominant reliance on deep learning models, with a notable emphasis on incorporating attention mechanisms to underscore relevant contextual information. Conversely, factorization machines and collaborative filtering techniques like tensor factorization still hold significance in managing high-dimensional contextual data. It is noteworthy that certain studies have explored the combination of these families of algorithms, aiming to achieve a more resilient and robust performance.

One of our main observations is related to the fact that the majority of context-aware recommenders depend on modeling techniques to incorporate context into the loop. This widespread preference is justified by its flexibility and effectiveness in enhancing performance, making it a robust paradigm for integrating context. Such approaches unveil concealed interactions between entities and contexts, pinpointing the relevant factors that contribute to the recommendation model's contextual understanding. Nevertheless, we have also noted instances where pre-filtering or post-filtering techniques persist. The underlying rationale for such approaches is to customize and simplify the complexity of the system by introducing more adaptable stages into the process, aligning with the specific requirements of the application domain.

Another important observation was that evaluation setups predominantly favored offline assessments, emphasizing ranking metrics over error-based ones. Yet, the need for incorporating metrics like novelty, diversity, and fairness is prominent, driven partly by the absence of standardized datasets and evaluation methodologies specific to context-aware recommender systems.

Domain-wise, context-aware recommender systems found substantial application in multimedia and e-commerce, followed by domains like news, food, and social networks. Conversely, domains such as books or e-documents were infrequently considered for context-aware solutions.

Finally, we highlighted the significance of experimental reproducibility, by underscoring the necessity for proper benchmarks to standardize the contextual spectrum.

Challenges and limitations

In our study, we identified a number of promising research problems for which few works exist so far and some inherent limitations of the existing literature:

- *Lack of formal definition of contextual features*: contextual characteristics may vary based on factors like data nature, acquisition, integration, and usage. For instance, determining the modeling assumptions regarding time as context poses questions about its interactional or representational nature. The same happens with precisely determining whether we encounter dynamism or observability features. On certain occasions, deciding the specific scenario to which a context belongs becomes a challenging task, highly dependent on the stage of the system being considered.
- *Novel and efficient methods*: we have identified some recommendation approaches that have been seldom used (e.g., reinforcement learning and bandits, topic modeling, novel neural approaches). While the universal and multidimensional nature of context may make applying any recommendation technique more difficult, the main challenge that should be addressed is that of efficiency, both in terms of memory and time, as contextual interactions are inherently more sparse than traditional non-contextual data.
- *Better and more complete evaluation*: more online or user studies are needed to properly validate CARS. One particular challenge those evaluations may face is how to adapt well-known protocols to contextual scenarios. Moreover, how to define beyond-accuracy evaluation metrics could be an interesting challenge with lots of

potential, even more if metrics are flexible and general enough to be applicable to any context domain.

- *Data heterogeneity*: as discussed throughout our work, data scarcity is one of the main limitations in current CARS research. Some researchers attempt to use synthetic data specifically to have full control of all the variables, but it is necessary to have real data with enough contextual attributes and values, and belonging to different domains, to extend and generalize current approaches beyond standard, small scenarios.
- *Reproducible experimental settings*: as it has happened in other areas, well-defined experimental protocols and tools are needed to obtain scientific progress. Currently, reproducibility in CARS is very limited, both in terms of baseline implementation, dataset variability, and evaluation frameworks. This challenge, when properly assessed and addressed, could spark a significant advancement in scientific integrity and trustworthiness. By establishing robust protocols for replication, the field can ensure that findings are reliable and verifiable. This, in turn, could lead to more accurate models and predictions, fostering a culture of transparency and accountability that is essential for the progress of research.

Future directions

Beyond addressing the aforementioned challenges, based on the performed review, we believe that future studies should significantly explore into a better understanding of the various aspects of contextual information (acquisition, observability, dynamism, and characterization) through providing transparent explanations. Future systems might integrate methods for producing explanations that humans can understand, elucidating the reasons behind particular recommendations through contextual reasoning. This area presents an opportunity for exploring and implementing LLM and Transformer-based models. However, several key challenges faced by LLMs need to be considered before being integrated in CARS. These challenges include hallucination (yielding grammatically correct but potentially incomplete or outdated responses based on statistical inference), limited knowledge, and ambiguous traceability, making it difficult to discern if answers are factual or mere guesses drawn from the training set. These issues can be mitigated through various strategies, such as incorporating knowledge graphs via ontologies, employing retrieval augmented generation (enhancing input provided to the LLM by sourcing data from databases), utilizing semantic search techniques, or integrating diverse data sources. Particularly in the tourism domain, these approaches have demonstrated efficacy and serve as pioneering examples from which LLMs can gather insights. Hence, an exploration on these aspects should extend to how effectively leverage such information for improved recommendations, while considering these challenges, in addition to the key task of clearly structuring an appropriate and unified framework to define the nature of the context.

Independently of having available a general context framework, another area deserving greater attention from the research community is context suggestion. Recommending appropriate contexts based on user preferences and item characteristics can not only improve the recommendation process but also enrich the explanations and personalization provided. Recommending the correct context at the right time can significantly enhance the user experience. Additionally, through context suggestion, a predetermined list of contexts can be prioritized and suggested to users, allowing for the inference of their preferences across different contexts based on their interactions. This can also aid users in making informed decisions.

Moreover, the pursuit of concepts such as explainability, diversity, and fairness through models utilizing causal inference represents a promising research avenue. It is essential for uncovering causal relationships and subsequently utilizing the acquired causal insights to improve recommendation systems.

Looking ahead, the future trajectory in this field should explore techniques that expand the contextual space, leveraging diverse data sources while considering scalability. By considering the last developments in this area applied to research projects in the field, the use of virtual reality in some of the domains discussed in our review, such as tourism or cultural heritage, would be a worth direction for the future; in particular, to integrate context modeling and personalization techniques within such technology, something that has been explored by other researchers, as in [143], but without the contextual dimension. Moreover, on the same domain of cultural heritage and applied to a European project called SPICE[4], the concept of *citizen curation* has been recently coined to allow citizens use their interpretations to support social cohesion across groups [144]. In fact, as already discussed from a broader perspective, such scenarios would benefit greatly from the use of contextual recommendation approaches, as they would be able to exploit social, technical, and geographical contexts of the citizens, key dimensions recognized in the literature to define the problem space [145]. Furthermore, a shift towards task-specific context-aware recommender systems is anticipated to sustain the vibrancy of this field in the coming years.

Supplementary information. The online version contains supplementary material.

Acknowledgments. This work has been supported by grant PID2022-139131NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe". The authors thank the reviewers for their thoughtful comments and suggestions.

Declarations

- Conflict of interest: the authors declare no competing interests.
- Ethics approval: not applicable.
- Consent to participate: not applicable.
- Consent for publication: not applicable.
- Availability of data and materials: the data used in this review is provided as supplementary information.
- Code availability: not applicable (no code was needed).
- Authors' contributions: both authors contributed equally. PM and AB: Conceptualization, Formal analysis, Investigation, Visualization, Writing—original draft, Writing—review and editing.

References

[1] Koren, Y., Rendle, S., Bell, R.M.: Advances in collaborative filtering. In: Ricci,

[4]Acronym for *Social cohesion, Participation, and Inclusion through Cultural Engagement*, https://spice-h2020.eu/.

F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, pp. 91–142. Springer, US (2022).

[2] Musto, C., de Gemmis, M., Lops, P., Narducci, F., Semeraro, G.: Semantics and content-based recommendations. In: Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, pp. 251–298. Springer, US (2022).

[3] Burke, R.D.: Hybrid recommender systems: Survey and experiments. User Model. User Adapt. Interact. **12**(4), 331–370 (2002)

[4] Abowd, G.D., Dey, A.K., Brown, P.J., Davies, N., Smith, M., Steggles, P.: Towards a better understanding of context and context-awareness. In: Handheld and Ubiquitous Computing, First International Symposium, HUC'99, Karlsruhe, Germany, September 27-29, 1999, Proceedings. Lecture Notes in Computer Science, vol. 1707, pp. 304–307. Springer, US (1999).

[5] Massimo, D., Ricci, F.: Harnessing a generalised user behaviour model for next-POI recommendation. In: Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018, pp. 402–406. ACM, US (2018).

[6] Smirnova, E., Vasile, F.: Contextual sequence modeling for recommendation with recurrent neural networks. In: Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems, DLRS@RecSys 2017, Como, Italy, August 27, 2017, pp. 2–9. ACM, US (2017).

[7] Adomavicius, G., Mobasher, B., Ricci, F., Tuzhilin, A.: Context-aware recommender systems. AI Mag. **32**(3), 67–80 (2011)

[8] Suhaim, A.B., Berri, J.: Context-aware recommender systems for social networks: Review, challenges and opportunities. IEEE Access **9**, 57440–57463 (2021)

[9] Casillo, M., Colace, F., Conte, D., Lombardi, M., Santaniello, D., Valentino, C.: Context-aware recommender systems and cultural heritage: a survey. J. Ambient Intell. Humaniz. Comput. **14**(4), 3109–3127 (2023)

[10] Bogina, V., Kuflik, T., Jannach, D., Bieliková, M., Kompan, M., Trattner, C.: Considering temporal aspects in recommender systems: a survey. User Model. User Adapt. Interact. **33**(1), 81–119 (2023)

[11] Kulkarni, S., Rodd, S.F.: Context aware recommendation systems: A review of the state of the art techniques. Comput. Sci. Rev. **37**, 100255 (2020)

[12] Le, Q., Vu, S., Nguyen, T., Le, T.: A state-of-the-art survey on context-aware recommender systems and applications. Int. J. Knowl. Syst. Sci. **12**(3), 1–20 (2021)

[13] Villegas, N.M., Sánchez, C., Díaz-Cely, J., Tamura, G.: Characterizing context-aware recommender systems: A systematic literature review. Knowl. Based Syst. **140**, 173–200 (2018)

[14] Raza, S., Ding, C.: Progress in context-aware recommender systems - an overview. Comput. Sci. Rev. **31**, 84–97 (2019)

[15] Villegas, N.M., Müller, H.A.: Managing dynamic context to optimize smart interactions and services. In: Chignell, M.H., Cordy, J.R., Ng, J., Yesha, Y. (eds.) The Smart Internet - Current Research and Future Applications. Lecture Notes in Computer Science, vol. 6400, pp. 289–318. Springer, US (2010).

[16] Kitchenham, B.A., Brereton, P., Budgen, D., Turner, M., Bailey, J., Linkman, S.G.: Systematic literature reviews in software engineering - A systematic literature review. Inf. Softw. Technol. **51**(1), 7–15 (2009)

[17] Karatzoglou, A., Amatriain, X., Baltrunas, L., Oliver, N.: Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In: Proceedings of the 2010 ACM Conference on Recommender Systems, RecSys 2010, Barcelona, Spain, September 26-30, 2010, pp. 79–86. ACM, US (2010).

[18] Rendle, S., Gantner, Z., Freudenthaler, C., Schmidt-Thieme, L.: Fast context-aware recommendations with factorization machines. In: Proceeding of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2011, Beijing, China, July 25-29, 2011, pp. 635–644. ACM, US (2011).

[19] Musto, C., Semeraro, G., Lops, P., de Gemmis, M.: Contextual eVSM: A content-based context-aware recommendation framework based on distributional semantics. In: E-Commerce and Web Technologies - 14th International Conference, EC-Web 2013, Prague, Czech Republic, August 27-28, 2013. Proceedings. Lecture Notes in Business Information Processing, vol. 152, pp. 125–136. Springer, US (2013).

[20] Wu, H., Yue, K., Liu, X., Pei, Y., Li, B.: Context-aware recommendation via graph-based contextual modeling and postfiltering. Int. J. Distributed Sens. Networks **11**, 613612–161361210 (2015)

[21] Unger, M., Bar, A., Shapira, B., Rokach, L.: Towards latent context-aware recommendation systems. Knowl. Based Syst. **104**, 165–178 (2016)

[22] Twardowski, B., Ryzko, D.: IoT and context-aware mobile recommendations using multi-agent systems. In: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT 2015, Singapore, December 6-9, 2015 - Volume I, pp. 33–40. IEEE Computer Society, US (2015).

[23] Hidasi, B., Tikk, D.: General factorization framework for context-aware recommendations. Data Min. Knowl. Discov. **30**(2), 342–371 (2016)

[24] Zheng, L., Cao, B., Noroozi, V., Yu, P.S., Ma, N.: Hierarchical collaborative embedding for context-aware recommendations. In: 2017 IEEE International Conference on Big Data (IEEE BigData 2017), Boston, MA, USA, December 11-14, 2017, pp. 867–876. IEEE Computer Society, US (2017).

[25] Costa, F.S.D., Dolog, P.: Collective embedding for neural context-aware recommender systems. In: Proceedings of the 13th ACM Conference on Recommender Systems, RecSys 2019, Copenhagen, Denmark, September 16-20, 2019, pp. 201–209. ACM, US (2019).

[26] Pan, H., Zhang, Z.: Research on context-awareness mobile tourism e-commerce personalized recommendation model. J. Signal Process. Syst. **93**(2-3), 147–154 (2021)

[27] del Carmen Rodríguez-Hernández, M., Ilarri, S.: AI-based mobile context-aware recommender systems from an information management perspective: Progress and directions. Knowl. Based Syst. **215**, 106740 (2021)

[28] del Carmen Rodríguez-Hernández, M., Ilarri, S.: Towards a context-aware mobile recommendation architecture. In: Awan, I., Younas, M., Franch, X., Quer, C. (eds.) Mobile Web Information Systems - 11th International Conference, MobiWIS 2014, Barcelona, Spain, August 27-29, 2014. Proceedings. Lecture Notes in Computer Science, vol. 8640, pp. 56–70. Springer, Germany (2014).

[29] Wu, J., He, X., Wang, X., Wang, Q., Chen, W., Lian, J., Xie, X.: Graph convolution machine for context-aware recommender system. Frontiers Comput. Sci. **16**(6), 166614 (2022)

[30] Haruna, K., Musa, A., Yunusa, Z., Ibrahim, Y., Rabiu, N.B., *et al.*: Location-aware recommender system: a review of application domains and current developmental processes. Science in Information Technology Letters **2**(1), 28–42 (2021)

[31] del Carmen Rodríguez-Hernández, M., Ilarri, S., Lado, R.T., Hermoso, R.: Location-aware recommendation systems: Where we are and where we recommend to go. In: Bouros, P., Lathia, N., Renz, M., Ricci, F., Sacharidis, D. (eds.) Proceedings of the Workshop on Location-Aware Recommendations, LocalRec 2015, Co-located with the 9th ACM Conference on Recommender Systems (RecSys 2015), Vienna, Austria, September 19, 2015. CEUR Workshop Proceedings, vol. 1405, pp. 1–8. CEUR-WS.org, US (2015).

[32] Hsieh, C., Yang, L., Wei, H., Naaman, M., Estrin, D.: Immersive recommendation: News and event recommendations using personal digital traces. In: Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016, pp. 51–62. ACM, US (2016).

[33] Sánchez-Moreno, D., Murciego, Á.L., Batista, V.F.L., Vicente, M.D.M., García, M.N.M.: Dynamic inference of user context through social tag embedding for music recommendation. In: Proceedings of the 15th ACM Conference on Recommender Systems, RecSys, Workshop on Context-Aware Recommender Systems, 2021, Amsterdam, Netherlands, 27th September-1st October , 2021. ACM, US (2021)

[34] Bauman, K., Tuzhilin, A.: Discovering contextual information from user reviews for recommendation purposes. In: Bogers, T., Koolen, M., Cantador,

I. (eds.) Proceedings of the 1st Workshop on New Trends in Content-based Recommender Systems Co-located with the 8th ACM Conference on Recommender Systems, CBRecSys@RecSys 2014, Foster City, Silicon Valley, California, USA, October 6, 2014. CEUR Workshop Proceedings, vol. 1245, pp. 2–9. CEUR-WS.org, US (2014).

[35] Zheng, Y., Mobasher, B., Burke, R.D.: Similarity-based context-aware recommendation. In: Wang, J., Cellary, W., Wang, D., Wang, H., Chen, S., Li, T., Zhang, Y. (eds.) Web Information Systems Engineering - WISE 2015 - 16th International Conference, Miami, FL, USA, November 1-3, 2015, Proceedings, Part I. Lecture Notes in Computer Science, vol. 9418, pp. 431–447. Springer, US (2015).

[36] Juan, Y., Zhuang, Y., Chin, W., Lin, C.: Field-aware factorization machines for CTR prediction. In: Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, USA, September 15-19, 2016, pp. 43–50. ACM, US (2016).

[37] Nguyen, T.V., Karatzoglou, A., Baltrunas, L.: Gaussian process factorization machines for context-aware recommendations. In: The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '14, Gold Coast , QLD, Australia - July 06 - 11, 2014, pp. 63–72. ACM, US (2014).

[38] Cheng, C., Xia, F., Zhang, T., King, I., Lyu, M.R.: Gradient boosting factorization machines. In: Eighth ACM Conference on Recommender Systems, RecSys '14, Foster City, Silicon Valley, CA, USA - October 06 - 10, 2014, pp. 265–272. ACM, US (2014).

[39] Sejwal, V.K., Abulaish, M., Jahiruddin: CRecSys: A context-based recommender system using collaborative filtering and LOD. IEEE Access **8**, 158432–158448 (2020)

[40] Yuan, F., Guo, G., Jose, J.M., Chen, L., Yu, H., Zhang, W.: LambdaFM: Learning optimal ranking with factorization machines using lambda surrogates. In: Proceedings of the 25th ACM International Conference on Information and Knowledge Management, CIKM 2016, Indianapolis, IN, USA, October 24-28, 2016, pp. 227–236. ACM, US (2016).

[41] Gu, Y., Song, J., Liu, W., Zou, L., Yao, Y.: Context aware matrix factorization for event recommendation in event-based social networks. In: 2016 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2016, Omaha, NE, USA, October 13-16, 2016, pp. 248–255. IEEE Computer Society, US (2016).

[42] Zheng, Y., Mobasher, B., Burke, R.D.: CSLIM: contextual SLIM recommendation algorithms. In: Eighth ACM Conference on Recommender Systems, RecSys '14, Foster City, Silicon Valley, CA, USA - October 06 - 10, 2014, pp. 301–304. ACM, US (2014).

[43] Zhao, G., Liu, Z., Chao, Y., Qian, X.: CAPER: context-aware personalized

emoji recommendation. IEEE Trans. Knowl. Data Eng. **33**(9), 3160–3172 (2021)

[44] Zheng, Y.: Context-aware mobile recommendation by A novel post-filtering approach. In: Brawner, K., Rus, V. (eds.) Proceedings of the Thirty-First International Florida Artificial Intelligence Research Society Conference, FLAIRS 2018, Melbourne, Florida, USA. May 21-23 2018, pp. 482–485. AAAI Press, US (2018).

[45] del Carmen Rodríguez-Hernández, M., Ilarri, S.: Pull-based recommendations in mobile environments. Comput. Stand. Interfaces **44**, 185–204 (2016)

[46] Zou, B., Li, C., Tan, L., Chen, H.: GPUTENSOR: efficient tensor factorization for context-aware recommendations. Inf. Sci. **299**, 159–177 (2015)

[47] Ge, H., Caverlee, J., Lu, H.: TAPER: A contextual tensor-based approach for personalized expert recommendation. In: Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, USA, September 15-19, 2016, pp. 261–268. ACM, US (2016).

[48] Zhao, J., Wang, W., Zhang, Z., Sun, Q., Huo, H., Qu, L., Zheng, S.: TrustTF: A tensor factorization model using user trust and implicit feedback for context-aware recommender systems. Knowl. Based Syst. **209**, 106434 (2020)

[49] Chen, H., Li, J.: Adversarial tensor factorization for context-aware recommendation. In: Proceedings of the 13th ACM Conference on Recommender Systems, RecSys 2019, Copenhagen, Denmark, September 16-20, 2019, pp. 363–367. ACM, US (2019).

[50] Shi, Y., Karatzoglou, A., Baltrunas, L., Larson, M.A., Hanjalic, A.: CARS2: learning context-aware representations for context-aware recommendations. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM 2014, Shanghai, China, November 3-7, 2014, pp. 291–300. ACM, US (2014).

[51] Jiang, Y., Bao, J., Ding, F.: A novel personalized recommendation model based on user's context. In: ICITEE-2019: 2nd International Conference on Information Technologies and Electrical Engineering, Zhuzhou, Hunan, China, December 6-7, 2019, pp. 154–11545. ACM, US (2019).

[52] Zheng, C., E, H., Song, M., Song, J.: CMPTF: contextual modeling probabilistic tensor factorization for recommender systems. Neurocomputing **205**, 141–151 (2016)

[53] Cunha, T., Soares, C., de Carvalho, A.C.P.L.F.: Metalearning for context-aware filtering: Selection of tensor factorization algorithms. In: Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys 2017, Como, Italy, August 27-31, 2017, pp. 14–22. ACM, US (2017).

[54] del Carmen Rodríguez-Hernández, M., Ilarri, S., Trillo, R., Hermoso, R.: Context-aware recommendations using mobile P2P. In: Pardede, E., Haghighi,

P.D., Salvadori, I.L., Steinbauer, M., Khalil, I., Anderst-Kotsis, G. (eds.) Proceedings of the 15th International Conference on Advances in Mobile Computing & Multimedia, MoMM 2017, Salzburg, Austria, December 4-6, 2017, pp. 82–91. ACM, US (2017).

[55] Wu, Q., Wang, H., Gu, Q., Wang, H.: Contextual bandits in a collaborative environment. In: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016, pp. 529–538. ACM, US (2016).

[56] Balakrishnan, A., Bouneffouf, D., Mattei, N., Rossi, F.: Using contextual bandits with behavioral constraints for constrained online movie recommendation. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden, pp. 5802–5804. ijcai.org, US (2018).

[57] Coviello, E., Ellis, K., Moerchen, F.: A scalable model for online contextual music recommendations. In: Proceedings of the Workshop on Context-Aware Recommender Systems (2021)

[58] Tang, L., Jiang, Y., Li, L., Li, T.: Ensemble contextual bandits for personalized recommendation. In: Eighth ACM Conference on Recommender Systems, RecSys '14, Foster City, Silicon Valley, CA, USA - October 06 - 10, 2014, pp. 73–80. ACM, US (2014).

[59] Tang, L., Jiang, Y., Li, L., Zeng, C., Li, T.: Personalized recommendation via parameter-free contextual bandits. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, August 9-13, 2015, pp. 323–332. ACM, US (2015).

[60] Zeng, C., Wang, Q., Mokhtari, S., Li, T.: Online context-aware recommendation with time varying multi-armed bandit. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016, pp. 2025–2034. ACM, US (2016).

[61] Unger, M., Shapira, B., Rokach, L., Bar, A.: Inferring contextual preferences using deep auto-encoding. In: Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, UMAP 2017, Bratislava, Slovakia, July 09 - 12, 2017, pp. 221–229. ACM, US (2017).

[62] Abinaya, S., Devi, M.K.K.: Enhancing top-N recommendation using stacked autoencoder in context-aware recommender system. Neural Process. Lett. **53**(3), 1865–1888 (2021)

[63] Abinaya, S., Alphonse, A.S., Abirami, S., Devi, M.K.K.: Enhancing context-aware recommendation using trust-based contextual attentive autoencoder. Neural Process. Lett. **55**(5), 6843–6864 (2023)

[64] Unger, M., Tuzhilin, A.: Hierarchical latent context representation for context-aware recommendations. IEEE Trans. Knowl. Data Eng. **34**(7), 3322–3334

(2022)

[65] Chen, K., Sun, S.: Cp-rec: Contextual prompting for conversational recommender systems. In: Williams, B., Chen, Y., Neville, J. (eds.) Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023, pp. 12635–12643. AAAI Press, US (2023).

[66] Prakash, T., Jalan, R., Singh, B., Onoe, N.: CR-SoRec: BERT driven consistency regularization for social recommendation. In: Zhang, J., Chen, L., Berkovsky, S., Zhang, M., Noia, T.D., Basilico, J., Pizzato, L., Song, Y. (eds.) Proceedings of the 17th ACM Conference on Recommender Systems, RecSys 2023, Singapore, Singapore, September 18-22, 2023, pp. 883–889. ACM, US (2023).

[67] Gu, T., Huang, K., Zhang, J., Zhang, K., Li, P.: Fast convolutional factorization machine with enhanced robustness. IEEE Trans. Knowl. Data Eng. **35**(3), 2579–2589 (2023)

[68] Kim, D.H., Park, C., Oh, J., Lee, S., Yu, H.: Convolutional matrix factorization for document context-aware recommendation. In: Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, USA, September 15-19, 2016, pp. 233–240. ACM, US (2016).

[69] Kristoffersen, M.S., Wieland, J.L., Shepstone, S.E., Tan, Z., Vinayagamoorthy, V.: Deep joint embeddings of context and content for recommendation. In: Proceedings of the 13th ACM Conference on Recommender Systems, RecSys, Workshop on Context-Aware Recommender Systems,2019, Copenhagen, Denmark, 20th September , 2019. ACM, US (2019)

[70] Hansen, C., Hansen, C., Maystre, L., Mehrotra, R., Brost, B., Tomasi, F., Lalmas, M.: Contextual and sequential user embeddings for large-scale music recommendation. In: RecSys 2020: Fourteenth ACM Conference on Recommender Systems, Virtual Event, Brazil, September 22-26, 2020, pp. 53–62. ACM, US (2020).

[71] Wang, D., Deng, S., Xu, G.: Sequence-based context-aware music recommendation. Inf. Retr. J. **21**(2-3), 230–252 (2018)

[72] Trichopoulos, G., Konstantakis, M., Alexandridis, G., Caridakis, G.: Large language models as recommendation systems in museums. Electronics **12**(18) (2023)

[73] Sattar, A., Bacciu, D.: Graph neural network for context-aware recommendation. Neural Process. Lett. **55**(5), 5357–5376 (2023)

[74] He, X., Liao, L., Zhang, H., Nie, L., Hu, X., Chua, T.: Neural collaborative filtering. In: Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017, pp. 173–182. ACM, US

(2017).

[75] Jawarneh, I.M.A., Bellavista, P., Corradi, A., Foschini, L., Montanari, R., Berrocal, J., Murillo, J.M.: A pre-filtering approach for incorporating contextual information into deep learning based recommender systems. IEEE Access **8**, 40485–40498 (2020)

[76] Unger, M., Tuzhilin, A., Livne, A.: Context-aware recommendations based on deep learning frameworks. ACM Trans. Manag. Inf. Syst. **11**(2), 8–1815 (2020)

[77] Manotumruksa, J., Macdonald, C., Ounis, I.: A contextual attention recurrent architecture for context-aware venue recommendation. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018, pp. 555–564. ACM, US (2018).

[78] Mei, L., Ren, P., Chen, Z., Nie, L., Ma, J., Nie, J.: An attentive interaction network for context-aware recommendations. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018, pp. 157–166. ACM, US (2018).

[79] Jhamb, Y., Ebesu, T., Fang, Y.: Attentive contextual denoising autoencoder for recommendation. In: Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR 2018, Tianjin, China, September 14-17, 2018, pp. 27–34. ACM, US (2018).

[80] Hu, B., Shi, C., Zhao, W.X., Yu, P.S.: Leveraging meta-path based context for top- N recommendation with A neural co-attention model. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018, pp. 1531–1540. ACM, US (2018).

[81] Wu, L., Chen, L., Hong, R., Fu, Y., Xie, X., Wang, M.: A hierarchical attention model for social contextual image recommendation. IEEE Trans. Knowl. Data Eng. **32**(10), 1854–1867 (2020)

[82] Chiang, J., Ma, C., Wang, C., Hao, P.: An adaptive, context-aware, and stacked attention network-based recommendation system to capture users' temporal preference. IEEE Trans. Knowl. Data Eng. **35**(4), 3404–3418 (2023)

[83] Tran, D.H., Sheng, Q.Z., Zhang, W.E., Tran, N.H., Khoa, N.L.D.: CupMar: A deep learning model for personalized news recommendation based on contextual user-profile and multi-aspect article representation. World Wide Web (WWW) **26**(2), 713–732 (2023)

[84] Xiao, J., Ye, H., He, X., Zhang, H., Wu, F., Chua, T.: Attentional factorization machines: Learning the weight of feature interactions via attention networks. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017, pp. 3119–3125. ijcai.org, US (2017).

[85] Chen, C., Zhang, M., Ma, W., Liu, Y., Ma, S.: Efficient non-sampling factorization machines for optimal context-aware recommendation. In: WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020, pp. 2400–2410. ACM / IW3C2, US (2020).

[86] Zheng, Y., Arias, G.F.: A family of neural contextual matrix factorization models for context-aware recommendations. In: UMAP '22: 30th ACM Conference on User Modeling, Adaptation and Personalization, Barcelona, Spain, July 4 - 7, 2022, Adjunct Proceedings, pp. 1–6. ACM, US (2022).

[87] Zhu, Q., Wang, S., Cheng, B., Sun, Q., Yang, F., Chang, R.N.: Context-aware group recommendation for point-of-interests. IEEE Access **6**, 12129–12144 (2018)

[88] Meng, Z., McCreadie, R., Macdonald, C., Ounis, I.: Variational bayesian representation learning for grocery recommendation. Inf. Retr. J. **24**(4-5), 347–369 (2021)

[89] Huang, W., Wu, Z., Liang, C., Mitra, P., Giles, C.L.: A neural probabilistic model for context based citation recommendation. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA, pp. 2404–2410. AAAI Press, US (2015).

[90] Beutel, A., Covington, P., Jain, S., Xu, C., Li, J., Gatto, V., Chi, E.H.: Latent cross: Making use of context in recurrent recommender systems. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM 2018, Marina Del Rey, CA, USA, February 5-9, 2018, pp. 46–54. ACM, US (2018).

[91] Gama, R., Fernandes, H.L.: An attentive RNN model for session-based and context-aware recommendations: a solution to the recsys challenge 2019. In: Proceedings of the Workshop on ACM Recommender Systems Challenge, Copenhagen, Denmark, September 2019, pp. 6–165. ACM, US (2019).

[92] Yuan, W., Wang, H., Yu, X., Liu, N., Li, Z.: Attention-based context-aware sequential recommendation model. Inf. Sci. **510**, 122–134 (2020)

[93] Ma, Y., Narayanaswamy, B.M., Lin, H., Ding, H.: Temporal-contextual recommendation in real-time. In: Gupta, R., Liu, Y., Tang, J., Prakash, B.A. (eds.) KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020, pp. 2291–2299. ACM, ??? (2020).

[94] Villa, A., Araujo, V., Cattan, F., Parra, D.: Interpretable contextual team-aware item recommendation: Application in multiplayer online battle arena games. In: Santos, R.L.T., Marinho, L.B., Daly, E.M., Chen, L., Falk, K., Koenigstein, N., de Moura, E.S. (eds.) RecSys 2020: Fourteenth ACM Conference on Recommender Systems, Virtual Event, Brazil, September 22-26, 2020, pp. 503–508. ACM, US (2020).

[95] Jankiewicz, P., Kyrashchuk, L., Sienkowski, P., Wójcik, M.: Boosting algorithms for a session-based, context-aware recommender system in an online travel domain. In: Proceedings of the Workshop on ACM Recommender Systems Challenge, Copenhagen, Denmark, September 2019, pp. 1–115. ACM, US (2019).

[96] Liu, H., Tong, Y., Zhang, P., Lu, X., Duan, J., Xiong, H.: Hydra: A personalized and context-aware multi-modal transportation recommendation system. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019, pp. 2314–2324. ACM, US (2019).

[97] Musto, C., Semeraro, G., Lops, P., de Gemmis, M.: Combining distributional semantics and entity linking for context-aware content-based recommendation. In: Dimitrova, V., Kuflik, T., Chin, D., Ricci, F., Dolog, P., Houben, G. (eds.) User Modeling, Adaptation, and Personalization - 22nd International Conference, UMAP 2014, Aalborg, Denmark, July 7-11, 2014. Proceedings. Lecture Notes in Computer Science, vol. 8538, pp. 381–392. Springer, Germany (2014).

[98] Lommatzsch, A.: Real-time news recommendation using context-aware ensembles. In: de Rijke, M., Kenter, T., de Vries, A.P., Zhai, C., de Jong, F., Radinsky, K., Hofmann, K. (eds.) Advances in Information Retrieval - 36th European Conference on IR Research, ECIR 2014, Amsterdam, The Netherlands, April 13-16, 2014. Proceedings. Lecture Notes in Computer Science, vol. 8416, pp. 51–62. Springer, Germany (2014).

[99] Chen, G., Chen, L.: Recommendation based on contextual opinions. In: Dimitrova, V., Kuflik, T., Chin, D., Ricci, F., Dolog, P., Houben, G. (eds.) User Modeling, Adaptation, and Personalization - 22nd International Conference, UMAP 2014, Aalborg, Denmark, July 7-11, 2014. Proceedings. Lecture Notes in Computer Science, vol. 8538, pp. 61–73. Springer, Germany (2014).

[100] Linda, S., Bharadwaj, K.: A genetic algorithm approach to context-aware recommendations based on spatio-temporal aspects. Integrated intelligent computing, communication and security, 59–70 (2019)

[101] Laß, C., Herzog, D., Wörndl, W.: Context-aware tourist trip recommendations. In: Neidhardt, J., Fesenmaier, D.R., Kuflik, T., Wörndl, W. (eds.) Proceedings of the 2nd Workshop on Recommenders in Tourism Co-located with 11th ACM Conference on Recommender Systems (RecSys 2017), Como, Italy, August 27, 2017. CEUR Workshop Proceedings, vol. 1906, pp. 18–25. CEUR-WS.org, US (2017).

[102] Aghdam, M.H., Hariri, N., Mobasher, B., Burke, R.D.: Adapting recommendations to contextual changes using hierarchical hidden markov models. In: Proceedings of the 9th ACM Conference on Recommender Systems, RecSys 2015, Vienna, Austria, September 16-20, 2015, pp. 241–244. ACM, US (2015).

[103] Gültekin, G., Bayat, O.: A Naïve Bayes prediction model on location-based recommendation by integrating multi-dimensional contextual information. Multim. Tools Appl. **81**(5), 6957–6978 (2022)

[104] Aliannejadi, M., Crestani, F.: Personalized context-aware point of interest recommendation. ACM Trans. Inf. Syst. **36**(4), 45–14528 (2018)

[105] Sejwal, V.K., Abulaish, M.: CAMO: A context-aware movie ontology generated from LOD and movie databases. Multim. Tools Appl. **80**(5), 7247–7269 (2021)

[106] Wang, M., Kawamura, T., Sei, Y., Nakagawa, H., Tahara, Y., Ohsuga, A.: Context-aware music recommendation with serendipity using semantic relations. In: Kim, W., Ding, Y., Kim, H. (eds.) Semantic Technology - Third Joint International Conference, JIST 2013, Seoul, South Korea, November 28-30, 2013, Revised Selected Papers. Lecture Notes in Computer Science, vol. 8388, pp. 17–32. Springer, ??? (2013).

[107] Karpus, A., Vagliano, I., Goczyla, K.: Serendipitous recommendations through ontology-based contextual pre-filtering. In: Kozielski, S., Mrozek, D., Kasprowski, P., Malysiak-Mrozek, B., Kostrzewa, D. (eds.) Beyond Databases, Architectures and Structures. Towards Efficient Solutions for Data Analysis and Knowledge Representation - 13th International Conference, BDAS 2017, Ustroń, Poland, May 30 - June 2, 2017, Proceedings. Communications in Computer and Information Science, vol. 716, pp. 246–259 (2017).

[108] Lommatzsch, A., Kille, B., Albayrak, S.: Incorporating context and trends in news recommender systems. In: Proceedings of the International Conference on Web Intelligence, Leipzig, Germany, August 23-26, 2017, pp. 1062–1068. ACM, US (2017).

[109] Bokde, D., Girase, S., Mukhopadhyay, D.: Matrix factorization model in collaborative filtering algorithms: A survey. Procedia Computer Science **49**, 136–146 (2015). Proceedings of 4th International Conference on Advances in Computing, Communication and Control (ICAC3'15)

[110] Rendle, S.: Factorization machines. In: ICDM 2010, The 10th IEEE International Conference on Data Mining, Sydney, Australia, 14-17 December 2010, pp. 995–1000. IEEE Computer Society, US (2010).

[111] Glowacka, D.: Bandit algorithms in recommender systems. In: Proceedings of the 13th ACM Conference on Recommender Systems, RecSys 2019, Copenhagen, Denmark, September 16-20, 2019, pp. 574–575. ACM, US (2019).

[112] Cheng, H., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Chai, W., Ispir, M., Anil, R., Haque, Z., Hong, L., Jain, V., Liu, X., Shah, H.: Wide & deep learning for recommender systems. In: Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, DLRS@RecSys 2016, Boston, MA, USA, September 15, 2016, pp. 7–10. ACM, US (2016).

[113] Guo, H., Tang, R., Ye, Y., Li, Z., He, X.: DeepFM: A factorization-machine based neural network for CTR prediction. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017, pp. 1725–1731. ijcai.org, US (2017).

[114] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Guyon, I., von Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R. (eds.) Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, pp. 5998–6008 (2017).

[115] Wang, D., Wan, F., Yu, D., Shen, Y., Xiang, Z., Xu, Y.: Context-and category-aware double self-attention model for next POI recommendation. Appl. Intell. **53**(15), 18355–18380 (2023)

[116] Selamat, A., Ibrahim, R., Alrashidi, M., Fujita, H.: Social recommender system based on cnn incorporating tagging and contextual features. J. Cases Inf. Technol. **26**(1), 1–20 (2024)

[117] Geng, S., Liu, S., Fu, Z., Ge, Y., Zhang, Y.: Recommendation as language processing (RLP): A unified pretrain, personalized prompt & predict paradigm (P5). In: Golbeck, J., Harper, F.M., Murdock, V., Ekstrand, M.D., Shapira, B., Basilico, J., Lundgaard, K.T., Oldridge, E. (eds.) RecSys '22: Sixteenth ACM Conference on Recommender Systems, Seattle, WA, USA, September 18 - 23, 2022, pp. 299–315. ACM, US (2022).

[118] Linda, S., Minz, S., Bharadwaj, K.K.: Effective context-aware recommendations based on context weighting using genetic algorithm and alleviating data sparsity. Appl. Artif. Intell. **34**(10), 730–753 (2020)

[119] Benouaret, I., Lenne, D.: Combining semantic and collaborative recommendations to generate personalized museum tours. In: Morzy, T., Valduriez, P., Bellatreche, L. (eds.) New Trends in Databases and Information Systems - ADBIS 2015 Short Papers and Workshops, BigDap, DCSA, GID, MEBIS, OAIS, SW4CH, WISARD, Poitiers, France, September 8-11, 2015. Proceedings. Communications in Computer and Information Science, vol. 539, pp. 477–487. Springer, US (2015).

[120] Colace, F., Santo, M.D., Lombardi, M., Mosca, R., Santaniello, D.: A multilayer approach for recommending contextual learning paths. J. Internet Serv. Inf. Secur. **10**(2), 91–102 (2020)

[121] Huang, W., Xu, R.Y.D.: Gaussian process latent variable model factorization for context-aware recommender systems. Pattern Recognit. Lett. **151**, 281–287 (2021)

[122] Adomavicius, G., Sankaranarayanan, R., Sen, S., Tuzhilin, A.: Incorporating contextual information in recommender systems using a multidimensional approach. ACM Trans. Inf. Syst. **23**(1), 103–145 (2005)

[123] Adomavicius, G., Tuzhilin, A.: Context-aware recommender systems. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 217–253. Springer, US (2011).

[124] Koren, Y., Bell, R.M., Volinsky, C.: Matrix factorization techniques for

recommender systems. Computer **42**(8), 30–37 (2009)

[125] Haruna, K., Akmar Ismail, M., Suhendroyono, S., Damiasih, D., Pierewan, A., Chiroma, H., Herawan, T.: Context-aware recommender system: A review of recent developmental process and future research direction. Applied Sciences **7**(12), 1211 (2017)

[126] Dacrema, M.F., Cremonesi, P., Jannach, D.: Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In: Bogers, T., Said, A., Brusilovsky, P., Tikk, D. (eds.) Proceedings of the 13th ACM Conference on Recommender Systems, RecSys 2019, Copenhagen, Denmark, September 16-20, 2019, pp. 101–109. ACM, US (2019).

[127] Kille, B., Lommatzsch, A.: Defining a meaningful baseline for news recommender systems. In: Özgöbek, Ö., Kille, B., Gulla, J.A., Lommatzsch, A. (eds.) Proceedings of the 7th International Workshop on News Recommendation and Analytics in Conjunction with 13th ACM Conference on Recommender Systems, INRA@RecSys 2019, Copenhagen, Denmark, September 20, 2019. CEUR Workshop Proceedings, vol. 2554, pp. 24–28. CEUR-WS.org, US (2019).

[128] Gunawardana, A., Shani, G., Yogev, S.: Evaluating recommender systems. In: Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, pp. 547–601. Springer, US (2022).

[129] Fang, Q., Xu, C., Hossain, M.S., Muhammad, G.: STCAPLRS: A spatial-temporal context-aware personalized location recommendation system. ACM Trans. Intell. Syst. Technol. **7**(4), 59–15930 (2016)

[130] McNee, S.M., Riedl, J., Konstan, J.A.: Being accurate is not enough: how accuracy metrics have hurt recommender systems. In: Extended Abstracts Proceedings of the 2006 Conference on Human Factors in Computing Systems, CHI 2006, Montréal, Québec, Canada, April 22-27, 2006, pp. 1097–1101. ACM, US (2006).

[131] Valcarce, D., Bellogín, A., Parapar, J., Castells, P.: Assessing ranking metrics in top-n recommendation. Inf. Retr. J. **23**(4), 411–448 (2020)

[132] Castells, P., Hurley, N., Vargas, S.: Novelty and diversity in recommender systems. In: Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, pp. 603–646. Springer, US (2022).

[133] Panniello, U., Tuzhilin, A., Gorgoglione, M.: Comparing context-aware recommender systems in terms of accuracy and diversity. User Model. User Adapt. Interact. **24**(1-2), 35–65 (2014)

[134] Adomavicius, G., Tuzhilin, A.: Context-aware recommender systems. In: Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, pp. 191–226. Springer, US (2015).

[135] Fu, Z., Yu, L., Niu, X.: TRACE: travel reinforcement recommendation based on location-aware context extraction. ACM Trans. Knowl. Discov. Data **16**(4),

65–16522 (2022)

[136] Lin, W., Zhang, X., Qi, L., Li, W., Li, S., Sheng, V.S., Nepal, S.: Location-aware service recommendations with privacy-preservation in the internet of things. IEEE Trans. Comput. Soc. Syst. **8**(1), 227–235 (2021)

[137] Braunhofer, M., Elahi, M., Ge, M., Ricci, F.: Context dependent preference acquisition with personality-based active learning in mobile recommender systems. In: Zaphiris, P., Ioannou, A. (eds.) Learning and Collaboration Technologies. Technology-Rich Environments for Learning and Collaboration - First International Conference, LCT 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part II. Lecture Notes in Computer Science, vol. 8524, pp. 105–116. Springer, Germany (2014).

[138] Bellogín, A., Said, A.: Improving accountability in recommender systems research through reproducibility. User Model. User Adapt. Interact. **31**(5), 941–977 (2021)

[139] Amigó, E., Deldjoo, Y., Mizzaro, S., Bellogín, A.: A unifying and general account of fairness measurement in recommender systems. Inf. Process. Manag. **60**(1), 103115 (2023)

[140] Ilarri, S., Lado, R.T., Hermoso, R.: Datasets for context-aware recommender systems: Current context and possible directions. In: 34th IEEE International Conference on Data Engineering Workshops, ICDE Workshops 2018, Paris, France, April 16-20, 2018, pp. 25–28. IEEE Computer Society, US (2018).

[141] del Carmen Rodríguez-Hernández, M., Ilarri, S., Hermoso, R., Lado, R.T.: DataGenCARS: A generator of synthetic data for the evaluation of context-aware recommendation systems. Pervasive Mob. Comput. **38**, 516–541 (2017)

[142] Xin, X., Chen, B., He, X., Wang, D., Ding, Y., Jose, J.M.: CFM: convolutional factorization machines for context-aware recommendation. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019, pp. 3926–3932. ijcai.org, US (2019).

[143] Bulla, L., Giorgis, S.D., Gangemi, A., Lucifora, C., Mongiovì, M.: Comparing user perspectives in a virtual reality cultural heritage environment. In: Indulska, M., Reinhartz-Berger, I., Cetina, C., Pastor, O. (eds.) Advanced Information Systems Engineering - 35th International Conference, CAiSE 2023, Zaragoza, Spain, June 12-16, 2023, Proceedings. Lecture Notes in Computer Science, vol. 13901, pp. 3–15. Springer, US (2023).

[144] Díaz-Agudo, B., Bosca, A., Bolioli, A., Diaz, G.J.J., Kuflik, T., Wecker, A.J.: Towards personalized social recommendations for cultural heritage activities: Methods and technology to enable cohesive and inclusive recommendations. In: Masthoff, J., Herder, E., Tintarev, N., Tkalcic, M. (eds.) Adjunct Publication of the 29th ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2021, Utrecht, The Netherlands, June 21-25, 2021, pp. 199–202. ACM, US (2021).

[145] Daga, E., Asprino, L., Damiano, R., Daquino, M., Díaz-Agudo, B., Gangemi, A., Kuflik, T., Lieto, A., Maguire, M., Marras, A.M., Pandiani, D.S.M., Mulholland, P., Peroni, S., Pescarin, S., Wecker, A.J.: Integrating citizen experiences in cultural heritage archives: Requirements, state of the art, and challenges. ACM Journal on Computing and Cultural Heritage **15**(1), 11–11135 (2022)