

A systematic literature review of multicriteria recommender systems

Original

A systematic literature review of multicriteria recommender systems / Monti, Diego; Rizzo, Giuseppe; Morisio, Maurizio. - In: ARTIFICIAL INTELLIGENCE REVIEW. - ISSN 0269-2821. - STAMPA. - (2020), pp. 1-42. [10.1007/s10462-020-09851-4]

Availability:

This version is available at: 11583/2834946 since: 2020-06-10T19:28:53Z

Publisher:

Springer

Published

DOI:10.1007/s10462-020-09851-4

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

Springer postprint/Author's Accepted Manuscript

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: <http://dx.doi.org/10.1007/s10462-020-09851-4>

(Article begins on next page)

A Systematic Literature Review of Multicriteria Recommender Systems

Diego Monti · Giuseppe Rizzo · Maurizio Morisio

Received: date / Accepted: date

Abstract Since the first years of the 90s, recommender systems have emerged as effective tools for automatically selecting items according to user preferences. Traditional recommenders rely on the relevance assessments that users express using a single rating for each item. However, some authors started to suggest that this approach could be limited, as we naturally tend to formulate different judgments according to multiple criteria. During the last decade, several studies introduced novel recommender systems capable of exploiting user preferences expressed over multiple criteria. This work proposes a systematic literature review in the field of multicriteria recommender systems. Following a replicable protocol, we selected a total number of 93 studies dealing with this topic. We subsequently analyzed them to provide an answer to five different research questions. We considered what are the most common research problems, recommendation approaches, data mining and machine learning algorithms mentioned in these studies. Furthermore, we investigated the domains of application, the exploited evaluation protocols, metrics and datasets, and the most promising suggestions for future works.

Keywords Recommender system · Multicriteria recommendation · Systematic literature review · Survey

Diego Monti
Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy
Tel.: +39-011-0907087
Fax: +39-011-0907099
E-mail: diego.monti@polito.it

Giuseppe Rizzo
LINKS Foundation, Via Pier Carlo Boggio 61, 10138 Turin, Italy
E-mail: giuseppe.rizzo@linksfoundation.com

Maurizio Morisio
Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy
E-mail: maurizio.morisio@polito.it

1 Background

Due to the large variety of products and digital contents available on the Web, an increasing number of people are interested in obtaining personalized suggestions, in order to reduce the effort of inspecting all the items in a catalog for selecting the best one according to their preferences. An automated tool capable of recommending items to users in a personalized way is defined as a Recommender System (RS) [20]. RSs were initially conceived at the beginning of the 90s [9] and, nowadays, they are considered part of a research field that is independent from Information Retrieval (IR) [12]. In fact, while search engines are based on queries and, therefore, they react to user stimuli, recommender systems try to automatically identify items that could be of interest for a certain user [4].

In general, recommender systems can be classified in different categories according to the recommendation approach. The most widespread categories of RSs are content-based, collaborative filtering, knowledge-based, and hybrid [2]. Content-based recommenders only rely on the past preferences of users in order to construct their profile and select suggested items. In contrast, collaborative filtering approaches analyze the behaviour of similar users for identifying candidate items. A knowledge-based recommender embeds domain-specific knowledge that is used for matching user requirements with items of potential interest. Finally, hybrid approaches combine in many different ways the previous methods. Other less common categories of recommender systems include community-based and demographic techniques [5]. A community-based recommender also considers the relationships of trust among its users, while a demographic RS mainly relies on demographic profiles.

Many authors already conducted different studies dealing with the topic of recommender systems. For example, Park et al. [17] reviewed 210 journal articles for analyzing the main application fields and data mining techniques exploited by recommenders from 2001 to 2010.

Various categories of recommender systems were investigated by previous surveys. Hong et al. [14] discussed the literature about context-aware recommenders, while Figueroa et al. [7] and Çano et al. [24] conducted systematic literature reviews about linked data and hybrid recommender systems respectively.

More recently, Quadrana et al. [19] classified different approaches according to their capability of managing sequences of items. Portugal et al. [18] reviewed commonly exploited machine learning techniques, while Zhang et al. [23] discussed the most promising deep learning methods for generating personalized items.

Traditionally, recommender systems ask users to quantify their preferences about items with a numerical value, called rating. Then, they try to compute ratings that users would assign to unknown items [2]. Those values represent the appropriateness of recommending an item to a certain user. In other words, recommender systems try to estimate a utility function $R(u, i)$ that, given a user u and an item i , predicts if i should be recommended to u .

However, this approach to the recommendation problem could be considered somewhat limited, as users typically judge items according to different criteria [1]. For example, we can easily imagine to assign different ratings to a movie, expressing how much we liked the story, the acting, the direction, and the visual effects. Such multiple ratings could be exploited by a recommender system in order to identify more effectively which items should be suggested. For this reason, different authors

started to propose multicriteria recommender systems, namely methods capable of suggesting items by relying on ratings provided over different criteria instead of a single one [2, 16].

In this work, we investigate the state of the art in the field of multicriteria recommender systems. We follow the systematic literature review protocol proposed by Kitchenham & Charters [15], in order to enable other researchers to easily verify and reproduce our work. We consider five different research questions that encompass various aspects of the reviewed studies.

In particular, we analyze the most important problems that multicriteria recommenders aim to address, as well as the exploited recommendation approaches, according to the taxonomy created by Burke [5]. We also describe the different machine learning and data mining techniques typically included in a multicriteria recommender and we identify which methods are frequently utilized in each recommendation phase, thus we try to describe the structure of an ideal multicriteria RS. We quantitatively measure the domains that are the most appropriate ones for such systems and we review how the proposed algorithms have been evaluated with respect to the experimental settings, the metrics, and the exploited datasets. Finally, we describe the most promising directions for future works that are mentioned in the reviewed studies.

We considered a total number of 93 studies, published from 2003 to 2018, to perform this systematic literature review. To the best of our knowledge, this is the first review conducted in the field of multicriteria recommender systems that follows a standardized and repeatable protocol. We aim that 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 as follows. In Section 2, we detail the protocol that we followed for conducting the review. Then, we present the quantitative results in Section 3 and we provide a possible interpretation of the outcomes of the review in Section 4. Finally, we conclude our work with Section 5, and, in Appendix A, we report the list of selected studies.

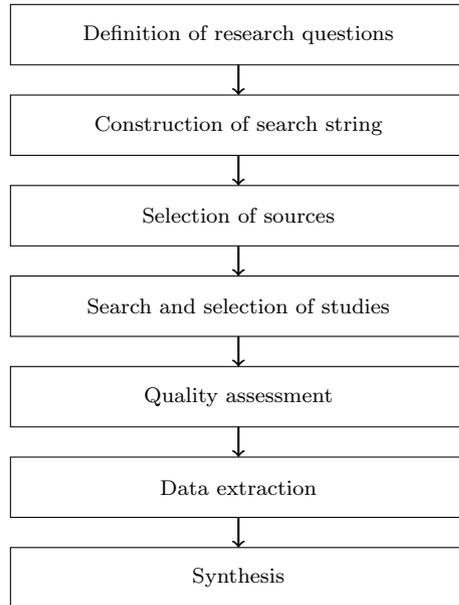
2 Methodology

We decided to perform this review according to the guidelines designed by Kitchenham & Charters for Systematic Literature Reviews (SLR) in the field of Software Engineering [15]. This method guarantees that the outcome of the review is verifiable and repeatable by other researches. The protocol, which is graphically illustrated in Figure 1, was developed by the first author and validated by the second and the third authors.

2.1 Research questions and search string

The purpose of this systematic literature review is to identify the studies describing multicriteria recommender systems and to understand the motivations behind their usage, the techniques employed, the experimental protocols used to validate them, and the related research challenges. For these reasons, we defined the following research questions.

Fig. 1: The systematic literature review protocol.



- RQ1 What are the most relevant studies addressing multicriteria RSs?
 RQ2 What are the most challenging problems faced by researchers?
 RQ3 What is the current state-of-the-art regarding multicriteria recommenders?
 RQ3.1 What are the approaches used by multicriteria recommenders?
 RQ3.2 Which techniques and methods have been proposed?
 RQ3.3 In which domains multicriteria recommender systems are applied?
 RQ4 How multicriteria recommenders are evaluated in literature?
 RQ4.1 Which protocols and frameworks are used for their evaluation?
 RQ4.2 Which metrics are considered during their evaluation?
 RQ4.3 Which datasets are used for testing the algorithms?
 RQ5 What are the most promising directions for future works?

In order to retrieve the studies related to multicriteria recommender systems, we defined the following preliminary set of keywords: {Multicriteria, Recommender System}. This initial set was expanded to include alternative spellings and we defined the search string used to query the digital sources as follows.

```
(multicriteria OR "multi criteria" OR "multi-criteria") AND
("recommender system" OR "recommendation system")
```

We selected six scientific digital libraries that contain primary studies related to the field of computer science, as detailed in Table 1. Other more general sources, like Google Scholar, were not included because they usually index studies already available in the primary sources.

Table 1: The sources considered during the search process.

Source	URL
ACM Digital Library	https://dl.acm.org
IEEE Xplore	http://ieeexplore.ieee.org
ISI Web of Knowledge	http://www.webofknowledge.com
ScienceDirect	https://www.sciencedirect.com
Scopus	https://www.scopus.com
Springer Link	https://link.springer.com

Table 2: The inclusion and exclusion criteria.

Code	Inclusion criteria
IC1	Papers describing multicriteria recommender systems
IC2	Papers published in conferences and journals
IC3	Papers written in English language
Code	Exclusion criteria
EC1	Papers not addressing recommender systems
EC2	Papers addressing RSs without multicriteria ratings
EC3	Papers that report only abstracts or posters
EC4	Papers that describe a planned research ¹
EC5	Grey literature and book chapters

2.2 Selection process

The selection process was performed during January 2019. We inserted the search query in the search field of the digital libraries selected as sources for the review and we retrieved all the studies identified by the respective search engines. Because of the high number of false positive results, we decided to limit our search to the title, abstract, and keywords with IEEE Xplore, ISI Web of Knowledge, and Scopus. The preliminary set initially contained 1256 studies. We checked their titles and authors in order to discover possible duplicates: after having removed duplicated results, the preliminary set was reduced to 950 studies.

Furthermore, for objectively identifying the studies to include in the review, we defined a set of inclusion and exclusion criteria, which are summarized in Table 2. We first applied the criteria in a coarse selection phase by only considering their abstracts and we obtained a list of 301 papers. Then, we analyzed again the available studies in a detailed selection phase by reading relevant portions of their content. We finally selected 93 studies as part of this literature review, as summarized in Table 3. The full list, sorted by source, year of publication and author, is available in Appendix A.

2.3 Quality assessment

For objectively assessing the quality of the studies selected as part of this review, we defined eight quality questions, as listed in Table 4. It is possible to assign

¹ We defined a planned research as a study that only contains a high level description of the proposed methodology, without the details necessary to implement it.

Table 3: The number of studies after each selection step.

Source	Search	Coarse	Detailed
ACM Digital Library	27	24	11
IEEE Xplore	38	36	15
ISI Web of Knowledge	31	25	6
ScienceDirect	344	81	19
Scopus	118	62	24
Springer Link	392	73	18
Total	950	301	93

Table 4: The quality questions.

Code	Quality question
QQ1	Did the study clearly describe the problems that it is addressing?
QQ2	Did the study review the related work for the problem?
QQ3	Did the study compare its approach with possible alternatives?
QQ4	Did the study describe the components of the proposed RS?
QQ5	Did the study provide an empirical evaluation of the solution?
QQ6	Did the study present a clear statement of the findings?
QQ7	Did the study analyze the application scenarios of the RS?
QQ8	Did the study recommend any further research activity?

to each question the scores of 0, 0.5, and 1, that correspond, respectively, to the answers *Yes*, *Partly*, and *No*. During the quality assessment phase, we provided an answer to each question for all studies included in the review. The scores were assigned by the first and second authors independently, then disagreeing values were discussed in order to reach a common decision. The third author checked a random subset of studies. We finally computed a global quality score for each paper by averaging the scores of all quality questions.

2.4 Data extraction

We carefully read multiple times the primary studies that are selected as part of this review. During this phase, we identified the data available in the works useful for providing an answer to the research questions introduced in Section 2.1. More in details, we looked for the information listed in Table 5. This process was supported by the data analysis software tool NVivo.² We relied on this tool to minimize the manual effort required for applying the methodology described in Section 2.5.

2.5 Synthesis

We synthesized the results of our review following the Cruzes & Dyba methodology [6] for combining and comparing the results of the primary studies that we considered. While reading the selected studies, we associated relevant portions of

² <https://www.qsrinternational.com/nvivo>

Table 5: The data extraction form.

Field	Description	RQ
Code	An internal identifier of the study	-
Title	-	RQ1
Authors	-	-
Publication year	-	RQ1
Publication name	-	-
Source	The digital library that contains the study	-
Type	Conference or journal	-
DOI	-	-
Research problem	The problem that the study tries to address	RQ2
Contribution	The description of the proposed method	RQ3.1
Implementation	How the method was implemented	RQ3.2
Domain	The domain of the recommended items	RQ3.3
Evaluation protocol	The protocol used to evaluate the method	RQ4.1
Evaluation metric	The metric used to compare the RS	RQ4.2
Dataset	The dataset used to execute the evaluation	RQ4.3
Limitation	The limitations of the proposed method	RQ5
Future work	The suggestions for future works	RQ5
Quality score	-	-

their text with codes. A code is a label applied to text segments that discuss the same theoretical or descriptive idea and that is used to aggregate in an organic way the data that we are analyzing. We initially defined some general codes associated with the research questions. Then, we created more specialized sub-codes related to the content of the studies, thus following an integrated approach that combines both inductive and deductive methods and that is considered the most appropriate one for a systematic review [6]. We subsequently aggregated the codes in themes, and we mapped these themes back to the original research questions. The outcomes of this last phase are reported in Section 3, grouped by research question.

3 Results

In this section, we highlight the findings of our systematic literature review regarding multicriteria recommender systems, according to the research questions introduced in Section 2.1. These results will be further discussed in Section 4.

3.1 Included studies

The main purpose of RQ1 is to identify the studies related to the topic of multicriteria recommender systems to be included in this review. Following the protocol detailed in Section 2, we identified a total number of 93 studies. These works have been presented during conferences or they have been published in scientific journals in a period of time from 2003 to 2018. In Figure 2, we detail the number of

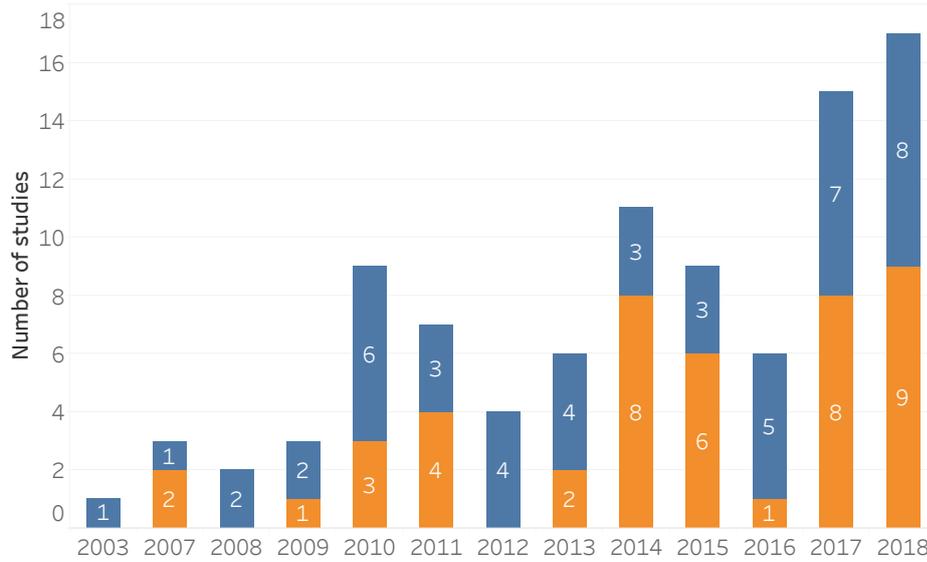


Fig. 2: This stacked barplot represents the number of included studies per year of publication. Blue studies have been presented in a conference, while orange studies have been published in a journal.

studies per year and per venue. It is possible to observe an increasing amount of studies published in the last years.³

We exploited the questions listed in Table 4 to assess the quality of the works included in this review. In Figure 3, we report the quality scores according to the publication venue. As expected, journal papers obtained, in general, higher scores with respect to conference papers. Furthermore, Figure 4 contains the average quality scores for each quality question. It is possible to observe that the highest scores are associated with QQ1 (*Did the study clearly describe the problems that it is addressing?*), while the lowest ones with QQ8 (*Did the study recommend any further research activity?*).

3.2 Research problems

In this section, we describe the main problems and challenges that multicriteria recommender systems aim to address and, thus, we provide an answer to RQ2. In total, we identified 10 different categories of problems that are mentioned in the reviewed studies. The number of studies for each category is summarized in Figure 5. It is important to observe that a single study may analyze different problems at the same time.

³ Please note that the results for the year 2019 are not available, as the selection was performed in January 2019.

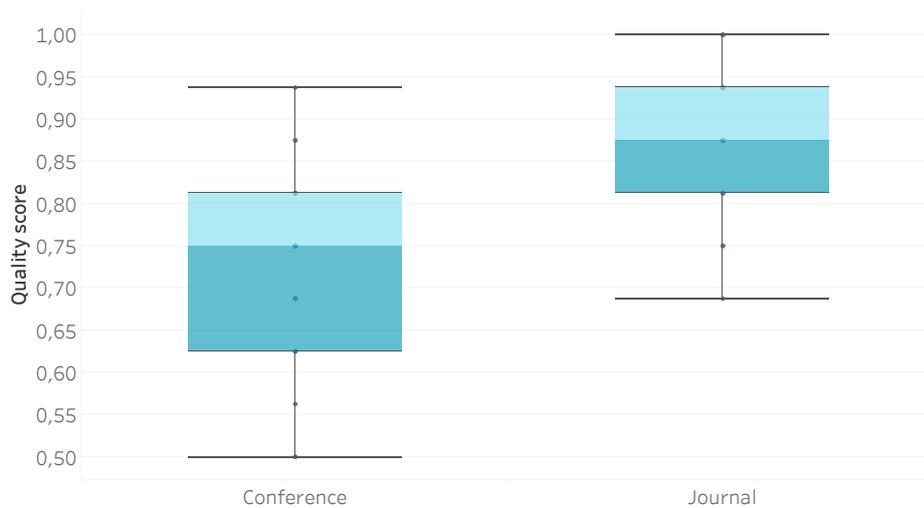


Fig. 3: This boxplot represents the distribution of quality scores per publication type. Studies published in journals have higher quality scores.

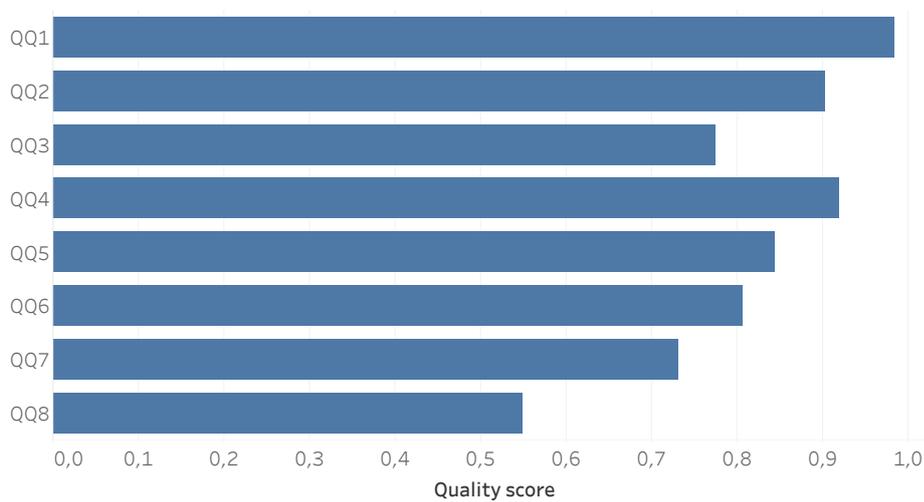


Fig. 4: This barplot represents the average quality score per quality question. The lowest quality score is associated with the description of future works.

3.2.1 Data sparsity

Data sparsity is the most frequent problem in this field and it is caused by the fact that users provide ratings for a limited number of items or criteria. While this is a well documented common issue of recommender systems, multicriteria user-item matrices may be even sparser, as they require more effort and time from the users of the system. In order to address this problem, several solutions are proposed in the reviewed studies. For example, the authors of [P3] suggest to

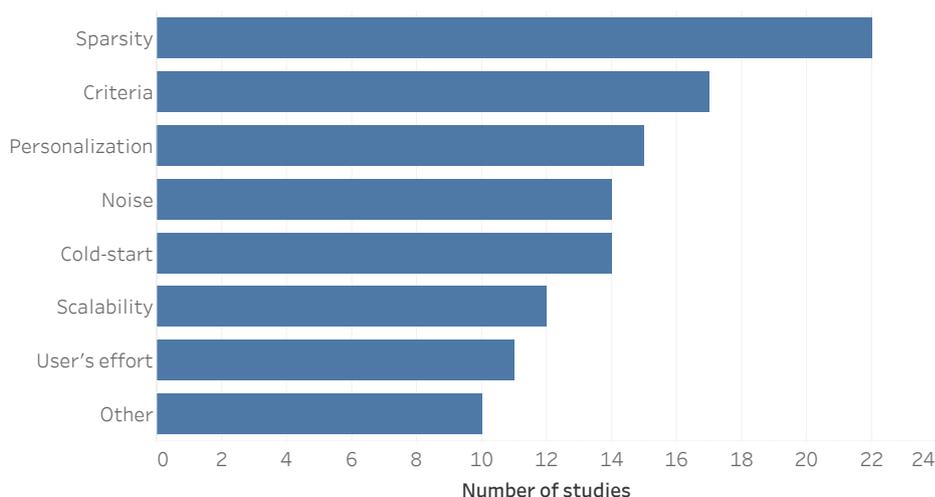


Fig. 5: This barplot represents the number of studies per research problem.

combine the multicriteria ratings using two different regression functions, one for the items and one for the users. The ratings estimated by the regression functions are then combined in order to minimize the prediction error. In [P18], a Bayesian latent model for multicriteria recommenders is exploited along with a support vector regression learner in order to mitigate the data sparsity problem. Another possible solution, as suggested in [P63], is represented by dimensionality reduction techniques, which can be used to obtain a more compact representation of each user. Finally, it is possible to integrate the available ratings with an external ontology [P14] or with a trust-based model [P72]. The data sparsity problem was identified in a total number of 22 studies.

3.2.2 Criteria weights

In order to provide accurate suggestions, it is of paramount importance being able to discover the relationships among the different criteria and to identify the most relevant ones for the target user. A wide range of possible solutions is available in the analyzed studies. For instance, the authors of [P8] identify the most important criteria for a user exploiting a statistical technique based on the average ratings of each item. Other studies analyze several machine learning methods. In [P9], the author proposes to consider *chains* of criteria instead of exploiting all criteria together: the rating on each criterion is estimated considering the previous predictions as context information. In [P5], the optimal weights are learned using particle swarm optimization, while in [P93] an artificial neural network is exploited for this purpose. Another popular solution is represented by decision making methods. As an example, in [P51] users are asked to perform pair-wise comparisons of the available criteria. The problem of selecting proper criteria weights was explicitly mentioned in 17 studies.

3.2.3 Personalization

Any recommender system should be capable of suggesting items that match the preferences of the target user. The possibility of exploiting multiple ratings for each item is often considered an effective way of increasing the accuracy of the recommendations, for example this fact is mentioned in [P15] and [P32]. It is also important being able to understand what are the most relevant criteria for each user, as suggested by the authors of [P8]. On the other hand, it is essential to avoid including criteria that are redundant, as this may negatively affect the performance of the system [P29]. In [P48], a multicriteria recommender system is combined with a content-based approach in order to generate better suggestions, while the authors of [P10] propose to increase the accuracy of the recommendations using a deep learning technique called Stacked Autoencoders. In general, this problem was explicitly considered by 15 studies.

3.2.4 Data noise

The presence of noise in data is typically related to the fact that users may provide ratings that are biased or even dishonest. For example, users may not understand the meaning of each criterion and may find difficult to express their preferences on a numerical scale, or may be bored by the request of assigning many ratings to a single item. In [P89], fuzzy logic techniques are exploited in order to address the uncertainty of user preferences, while in [P72] such techniques are combined with a trust-based model. The authors of [P3] propose to mitigate this problem by performing feature selection in a pre-processing step, where the most relevant criteria of a certain dataset are identified. Another possible solution to this problem is represented by the idea of considering the numerical differences between ratings instead of their absolute values in the recommendation process [P32]. Data noise was addressed by 14 studies.

3.2.5 Cold-start

The cold-start problem is a well-known issue in the field of recommender systems based on collaborative filtering approaches. It can be defined as the impossibility of creating reliable suggestions due to the lack of data regarding a new user or a new item. The authors of [P60] aim to solve it by providing non-personalized recommendations to new users and exploiting content-based features when a new item is added to the system. A different approach is represented by the elicitation of user preferences using decision making techniques and multicriteria ratings [P66]. In [P47], a multicriteria implicit feedback method based on user behavior analysis is discussed, while the authors of [P15] propose to tackle this issue with a trust-based model. As a last example, a knowledge-based method that is immune to the cold-start problem is illustrated in [P12]. Cold-start was considered a research issue in 14 studies.

3.2.6 Scalability

Scalability is a general problem of collaborative filtering recommender systems, especially for the ones developed in an academic context as proof-of-concept. Be-

cause many multicriteria recommenders require multiple runs of such algorithms, it is reasonable to suppose that scalability issues are even more widespread. For example, this issue is discussed by the authors of [P24], who describe a multicriteria recommender that exploits a distributed architecture based on Apache Spark. In [P32], a clustering algorithm is applied for creating groups of similar users that can be used to compute predictions in a scalable way. A popular dimensionality reduction technique discussed by [P19] and [P84] is the higher order single value decomposition. In total, 12 studies considered the scalability problem.

3.2.7 User's effort

As multicriteria recommenders usually require many ratings for each user and item pair, explicit elicitation methods are intrusive and may waste the user's effort and time. For this reason, the authors of [P30] and of [P50] propose a multicriteria recommender based only on implicit feedback. Another possible solution is described in [P63], where their authors develop a hybrid profiling framework for reusing traditional ratings with multicriteria recommenders. A similar approach to this problem is presented in [P48], where multicriteria ratings are computed starting from single ratings and content-based information. This problem was discussed by 11 studies.

3.2.8 Other research problems

Other research problems are mentioned in a more limited number of studies. In particular, 4 studies analyzed the issues related to the selection of a proper similarity metric in the context of multidimensional neighborhood-based collaborative filtering, while 3 studies reported the challenges related to the execution of a reliable evaluation protocol. Finally, 3 more recent studies mentioned the problem of fairness in the selection of the recommended items, both with respect to the unique peculiarities of the users and to the characteristics of the catalog.

3.3 Recommendation approaches

In order to provide an answer to RQ3.1, we analyzed the studies included in this systematic literature review and we classified them according to the taxonomy provided by Burke [5]. This taxonomy has become a widespread way of characterizing different recommendation approaches. However, we decided not to consider hybrid recommender systems, as almost all multicriteria recommenders would fall in this category. For this reason, if a study combines multiple approaches, it will be included in all the categories of the different methods that are mentioned in the study. We summarize the number of available studies for each recommendation approach in Table 6.

3.3.1 Collaborative filtering

Collaborative filtering is the most popular recommendation technique described in the reviewed literature. In a traditional recommender system, the users whose

Approach	Studies
Collaborative filtering	82
Content-based	16
Knowledge-based	12
Community-based	5
Demographic	4

Table 6: The number of studies per recommendation approach.

rating behaviour is similar to the one of the target user are exploited for selecting the items to be suggested. In a multicriteria recommender, a popular approach consists in applying collaborative filtering algorithms on the available criteria and then combining the results in a global estimated rating, like in [P13] and [P28]. An alternative to the user-based approach is represented by the item-based collaborative filtering, where the similarity is computed among the items, as discussed, for example, in [P65]. The authors of [P3] and [P45] describe how to combine user-based and item-based models in a comprehensive approach. Collaborative filtering may also be implemented with a model-based approach, for instance matrix factorization (e.g., in [P21] and [P75]). In general, this technique may be combined with other methods, like content-based approaches [P16], clustering [P32], and fuzzy logic [P60]. In total, collaborative filtering was exploited by 82 studies.

3.3.2 Content-based

A content-based recommender system considers the previous preferences of a user in order to build a profile and to select items with similar characteristics. For example, the authors of [P79] describe a recommender system of research papers that relies on content-based and multicriteria collaborative filtering algorithms. In [P43], user profiles are created from multicriteria ratings and, then, they are exploited for building clusters of similar users. The authors of [P30] suggest to identify the user's category-wise preferences for each criterion with a content-based approach. A possible source of structured information regarding items are external ontologies, as they are discussed in [P16]. In [P6] and [P20], user's reviews are mined to identify the most important features of each item. Content-based approaches were mentioned by 16 studies.

3.3.3 Knowledge-based

A knowledge-based recommender system relies on an externally encoded domain knowledge in order to match the user profiles with certain item features. For example, in [P69] a mobile recommender suggests restaurants considering their geographical location and cuisine. The author of [P46] describes a method for asking users to express their preferences regarding the features of a smartphone. In a similar vein, the recommender system presented in [P76] exploits user profiles and fuzzy set theory in order to suggest cities to be visited. In [P47], the authors of the study propose a set of rules for building a personalized list of recommended movies. Knowledge-based techniques were identified in 12 studies.

Technique	Studies
k -NN	35
Decision analysis	26
Fuzzy logic	16
Regression analysis	15
Clustering	15
Matrix manipulation	12
Neural networks	12
Genetic algorithms	7
Other	7

Table 7: The number of studies per recommendation technique.

3.3.4 Community-based

If a recommender system also considers the relations of friendship and trust among its users, it is defined as community-based. For example, the authors of [P72] propose a multicriteria collaborative filtering recommender that is enhanced by considering trust as an additional weight during the hybrid prediction phase. A similar approach is followed by [P61] and [P91], where the trust score for each user is computed only considering rating data and it is combined with the results of a collaborative filtering algorithm. Community-based recommenders were discussed by 5 studies.

3.3.5 Demographic

A demographic recommender considers the demographic profile of the user for suggesting items. For example, the authors of [P31] describe a multicriteria recommender for groups where users are clustered also according to their demographic profile. The other studies that mention demographic information are [P45], [P43], and [P11]. Demographic recommenders were considered by 4 studies.

3.4 Multicriteria techniques

In the following, we analyze the main techniques and methods related to multicriteria recommender systems that we identified in the reviewed studies and, therefore, we provide an answer to RQ3.2. In most studies, the recommenders combine different techniques, for example k -NN may be exploited to estimate unknown ratings, while decision analysis to merge the different ratings in a global prediction. We summarize the most frequent ones in Table 7.

3.4.1 k -NN

k -NN is a classification method used in data mining applications that relies on the similarity of the instances to be classified with the training examples. In the context of collaborative filtering, k -NN is exploited to find similar users or items considering their neighborhood. For example, in [P13], a k -NN collaborative filtering method is applied to each criterion separately, like in traditional recommenders.

In contrast, a different approach to this problem is considering all the criteria together when applying the k -NN algorithm. To this end, it is necessary to rely on a multidimensional distance, like the Manhattan, Euclidean, or Chebyshev distance, as described in [P53]. k -NN is the most popular data mining technique applied to multicriteria recommender systems, as it was identified in 35 studies.

3.4.2 Decision analysis

In order to rank items with contrasting criteria, it is possible to exploit the tools provided by multiple-criteria decision analysis, which is a sub-field of operations research. In general, different methods are available in order to support users in making complex decisions, and some of these methods have also been applied to multicriteria recommender systems. For example, in [P38], an analysis hierarchy process (AHP) is used in order to help users to evaluate the relative importance of each criterion. In [P59], the UTA* algorithm is exploited for constructing user profiles that are subsequently grouped according to their preferences. Other decision analysis methods mentioned in the reviewed studies are, for instance, ELECTRE [P37], SMART [P35], TOPSIS [P31], and UTADIS [P78]. In total, decision analysis techniques were identified in 26 studies.

3.4.3 Fuzzy logic

Fuzzy logic is a mathematical model that can be used to represent the concept of partial truth. This model is exploited to formalize the vagueness and uncertainty that are usually associated with user ratings. For example, in [P17], [P60], and [P25], ratings are expressed using linguistic terms in a qualitative way, considering that each term may have a different meaning according to the user. Furthermore, the authors of [P69] use the AHP decision analysis method in the fuzzy domain using fuzzy numbers instead of real numbers. In [P84], fuzzy rules that express how to build global ratings are identified for each cluster of users. Fuzzy logic was exploited as a recommendation technique in 16 studies.

3.4.4 Regression analysis

Regression analysis is a set of statistical techniques for predicting the value of a dependent variable given one or more independent variables. In the reviewed studies, such techniques are typically used to estimate the global rating of an item considering the predicted ratings for each criterion. For example, the authors of [P75] find the weights of the aggregation function with a linear regression model that is learned for each user. In [P88], a non-personalized linear regression model is first used to aggregate the similarities among users, and then to estimate the final ratings. A different approach is represented by Support Vector Regression (SVR): for instance, in [P82], a SVR model is trained for each user in order to synthesize the overall rating. We found regression analysis techniques in 15 studies.

3.4.5 Clustering

Clustering is an exploratory data mining approach that consists in grouping objects in cohesive sets. A typical application of such techniques to multicriteria

recommender systems is represented by the identification of users with similar profiles. For instance, in [P32], [P59], and [P52], the global K-means clustering algorithm is exploited in order to create groups of users with similar preferences. In [P82], clusters of users are created according to the importance given to each criterion. On the other hand, the authors of [P56] propose to cluster the items and, then, to learn an aggregation function for each user and item cluster. A clustering technique is also exploited to identify malicious users in the context of robust recommenders [P27]. In total, clustering algorithms were mentioned in 15 studies.

3.4.6 Matrix manipulation

In the reviewed studies, we identified different techniques used to compute predicted ratings with mathematical operations on matrices. For example, in [P90], the Singular Value Decomposition (SVD) method is exploited to compute unknown ratings for each criterion. In contrast, the authors of [P19] propose to reduce the dimensionality of the user, item, and criterion tensor with the Higher-Order Singular Value Decomposition (HOSVD) method and then to apply a collaborative filtering algorithm to the resulting matrix. In [P64], a matrix factorization technique is applied to a utility matrix estimated from the multicriteria ratings using a neural network model trained considering each user. A different approach is followed by the authors of [P92], which proposes a factorization machine model for representing all multicriteria ratings together. Matrix manipulation techniques were discussed in 12 studies.

3.4.7 Neural networks

Neural networks are usually applied to multicriteria recommender systems in order to aggregate the predicted ratings for each criterion in a global score. For example, in [P24], a single layer PERCEPTRON algorithm is selected for this task, while the authors of [P73] propose a neural network trained with the simulated annealing algorithm. In [P84], an Adaptive Neuro-Fuzzy Inference System (ANFIS) is exploited for extracting fuzzy rules for each cluster of users; such rules are later applied to predict the overall rating. A different approach is described in [P93], where a neural factorization machine is used to model the interactions among users, items, and criteria, and in [P37], where a single layer PERCEPTRON is exploited to estimate the similarity among users. We identified neural network approaches in 12 studies.

3.4.8 Genetic algorithms

Genetic algorithms can be considered a family of optimization techniques and, in the reviewed studies, they are typically used to determine the weights of each criterion. For example, in [P55] and [P87], a genetic algorithm is run for each user in order to construct a personalized aggregation function. In contrast, the authors of [P3] propose to use it for performing a feature selection of the available criteria in order to identify an optimal set of dimensions. In total, genetic algorithms were exploited by 7 studies.

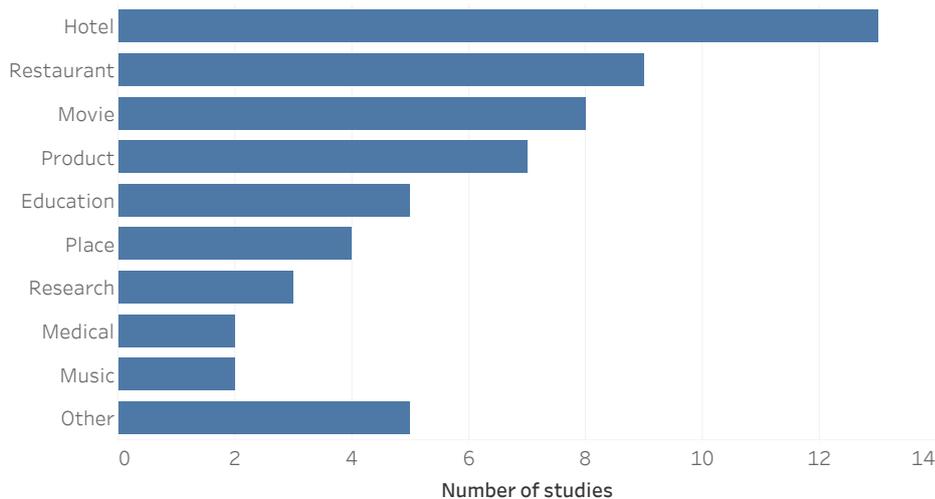


Fig. 6: This barplot represents the number of studies per application domain.

3.4.9 Other techniques

Other less frequent techniques described in the reviewed studies include statistical modeling [P18], particle swarm optimization [P5], and natural language processing [P36]. Such techniques were found in 7 studies.

3.5 Application domains

We analyzed the application domains of the multicriteria recommender systems described in the reviewed studies in order to provide an answer to RQ3.3. A graphical summary listing the categories of recommended items, considering possible examples and the experimental evaluation, is available in Figure 6. We observe that the majority of studies propose to apply multicriteria recommenders to domains related to tourism and travel. For example, 13 studies describe recommenders for hotels, 9 related to restaurants, and 4 dealing with tourist places. Another popular domain is related to movies, mentioned in 8 studies. Other domains include consumer electronics products and education, described in 7 and 5 studies respectively. Research papers were discussed in 3 studies, while medical treatments and music in 2 studies each. Less popular domains, identified only in 1 study and grouped in a miscellaneous category, are business and romantic partners, investment solutions, electronic books, and job opportunities.

Different categories of criteria are selected by researchers according to the domain. For example, popular criteria for hotels are *rooms*, *location*, *cleanliness*, *service*; for restaurants *food quality*, *service*, *presentation*, *taste*; for tourist places *architectural style*, *ease of access* and *welcome quality*; for movies *story*, *acting*, *direction* and *visuals*; for consumer electronics products *type*, *brand*, *weight*, *size*; for learning resources *subject relevance* and *educational value*.

Evaluation protocol	Studies
Offline comparison	74
User study	12
Case study	4
No evaluation	3

Table 8: The number of studies for each evaluation protocol.

3.6 Evaluation protocols

In this section, we describe the evaluation protocols followed by the reviewed studies, in line with RQ4.1. We grouped the possible evaluation strategies in four main categories, which are summarized in Table 8.

3.6.1 *Offline comparison*

We discovered that 74 studies compare the proposed solution with other approaches using an offline evaluation. Multicriteria recommender systems are usually compared against traditional baselines such as single-criteria recommender or weighted average multicriteria approaches. Different studies consider the most similar methods already available in literature, while few studies only compare the described technique with itself, analyzing several configuration parameters.

3.6.2 *User study*

A different approach is represented by the execution of user studies, which were carried out in 12 works. For example, the authors of [P47] created a movie recommender system that was tested by 567 users. The researchers computed different metrics considering their behaviour while utilizing the recommender. In contrast, in [P86] 158 users were asked to fill out a questionnaire in order to compare different recommendation models. User studies are also exploited to evaluate the usability of the system, as done, for instance, in [P58].

3.6.3 *Case study*

We also identified 4 studies that evaluated the proposed approach by describing a case study. For example, in [P4], a possible application of a multicriteria recommender to the movie domain is discussed, while the authors of [P54] empirically compare the suggested restaurants considering different user profiles.

3.6.4 *No evaluation*

Finally, 3 studies performed no evaluation of the multicriteria recommender presented in the paper. For example, in [P15], the evaluation of the proposed model is left as a future work.

Evaluation metric	Studies
Rating accuracy	59
Usage accuracy	43
Coverage	11
Ranking accuracy	10
Scalability	8
Other	5

Table 9: The number of studies for each evaluation metric.

3.7 Evaluation metrics

In the following, we discuss the metrics exploited in the reviewed studies for conducting the experimental evaluation of the proposed solutions in order to answer to RQ4.2. We decided to classify them according to the dimensions related to the recommender system proprieties described by Gunawardana et al. [10]. The identified category for each evaluation metric and the associated number of studies are reported in Table 9.

3.7.1 Rating accuracy

This category includes metrics designed to evaluate the capability of the system to correctly estimate user ratings. In particular, 51 studies report the Mean Absolute Error (MAE), 20 the Root Mean Squared Error (RMSE), and 3 the Mean Squared Error (MSE). Other metrics exploited by 1 study each are the coefficient of determination (R^2) and the Mean Absolute Percentage Error (MAPE). In total, rating accuracy metrics are mentioned in 59 studies.

3.7.2 Usage accuracy

If the goal of a recommender is to predict a list of items, it is possible to evaluate the usage accuracy of the available suggestions. Precision is considered in 36 studies, recall in 28, F1 in 23 studies, and Area Under the Curve (AUC) in 4 studies. Usage accuracy is the second most popular category, as it was identified in 43 studies.

3.7.3 Coverage

The metric of coverage was computed in 11 studies. Even if this metric can be evaluated both at the level of users and at the level of items, all the studies included in this review considered the coverage of the item space, also known as catalog coverage [13].

3.7.4 Ranking accuracy

The correctness of the ranking in the recommended lists of items was analyzed by 10 studies. In details, 8 studies exploit the Normalized Discounted Cumulative Gain (nDCG) metric, while 4 studies the Fraction of Concordant Pairs (FCP). A less popular metric, described by 1 study, is the Kendall's τ .

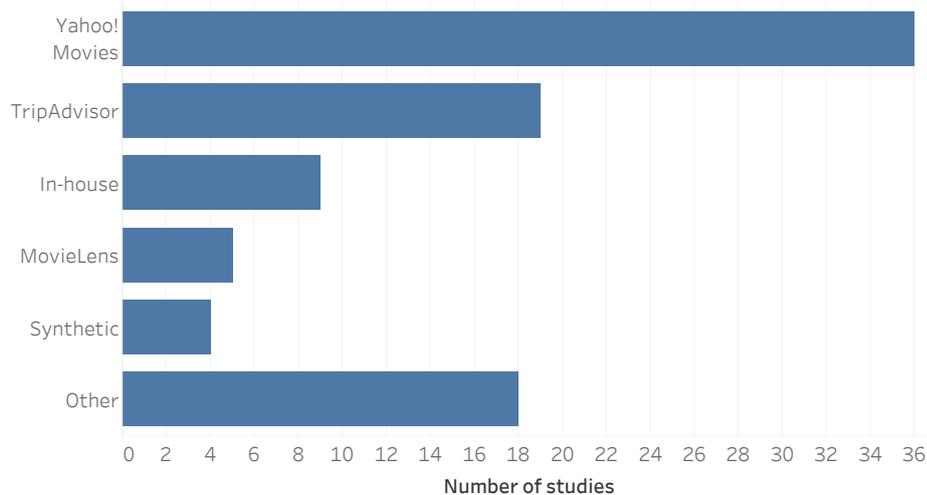


Fig. 7: This barplot represents the number of studies per evaluation dataset.

3.7.5 Scalability

The authors of 8 studies evaluated the scalability of the proposed approach. A typical metric used to this purpose is the time required to compute the predictions, which is reported by 6 studies. In contrast, 2 studies exploit the speed of the recommendations.

3.7.6 Other metrics

Additional metrics identified in the reviewed studies include the utility of the suggested items and the system satisfaction, evaluated with a user study, and the robustness of the recommendations. Finally, the authors of [P68] defined a combined metric.

3.8 Evaluation datasets

In line with RQ4.3, we analyzed the datasets exploited for conducting the experimental evaluation of the techniques described in the reviewed studies. In total, 74 studies mentioned at least one dataset: this result is consistent with the number of studies that performed an offline comparison, as reported in Section 3.6. We summarize the studies for each dataset in Figure 7.

3.8.1 Yahoo! Movies

Yahoo! Movies was a website, part of the Yahoo! network, that provided information and reviews about movies. Among other features, users were able to rate each movie considering five criteria: story, acting, direction, visuals, and overall. Yahoo!

Research provides a public Yahoo! Movies dataset, but it does not include multicriteria ratings.⁴ To address this issue, the authors of [P3] created a multicriteria version of the same dataset by crawling the Yahoo! Movies website. In total, 36 studies exploit the Yahoo! Movies dataset, either the version obtained by Jannach et al. [P3] or other versions crawled by different researchers, such as [P18], [P29], and [P53].

3.8.2 *TripAdvisor*

TripAdvisor is a website that contains restaurant and hotel reviews. Similarly to Yahoo! Movies, there is no official multicriteria rating dataset, as different researchers crawled the website and created their own version, typically exploiting hotel ratings and reviews. For example, this approach was followed by the authors of [P14] and [P83]. The TripAdvisor dataset collected by the authors of [P1] is publicly available.⁵ Also the TripAdvisor dataset created by Wang et al. [22] and used in [P91] is available online.⁶ In total, the TripAdvisor dataset was mentioned by 19 studies.

3.8.3 *In-house*

We identified 9 studies that created a multicriteria dataset in-house for conducting an offline comparison. For example, the authors of [P60] collected 9,628 ratings about songs that were later used to evaluate a music recommender system. In [P19], different students were invited to provide ratings about universities.

3.8.4 *MovieLens*

Even if the MovieLens datasets only contain single criteria ratings, they were also exploited for evaluating multicriteria recommender systems. For instance, in [P2] and in [P61], MovieLens 100K was transformed in a four criteria dataset. A similar approach was followed in [P48] with MovieLens 10M, where a method capable of extracting multicriteria preferences from traditional ratings using external aggregate ratings and descriptive data is discussed. In total, the MovieLens datasets were mentioned by 5 studies.

3.8.5 *Synthetic*

Because of the lack of public multicriteria datasets, some researchers created synthetic ratings in order to evaluate their approach. For example, the authors of [P78] simulated a dataset about equity fund recommendations. In [P77], a testing tool named CollaFiS, capable of building multicriteria datasets, is discussed. This approach was followed by 4 studies.

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

⁵ <https://www.cs.cmu.edu/~jiweil/html/hotel-review.html>

⁶ <http://www.cs.virginia.edu/~hw5x/Data/LARA/TripAdvisor>

Future work	Studies
Extend the solution	33
Include more data or criteria	24
Improve the evaluation	14
Identify significant criteria	13
Increase the scalability	10
Consider different domains	7
Other	9

Table 10: The number of studies per category of future work.

3.8.6 Other datasets

Less common datasets, exploited by 1 or 2 studies each, include, for example, Bcn Restaurantes [P39], HRS.com [P91], and RateBeer [P93]. Such various datasets were considered by 18 studies.

3.9 Future works

In the following, we analyze the suggestions for future works mentioned in the reviewed studies in order to provide an answer to RQ5. A summary of our findings is available in Table 10.

3.9.1 Extend the solution

Different authors propose to extend or modify the described recommender system for increasing its accuracy. This future work is related to the problem of personalization, discussed in Section 3.2.3. Common suggestions include adding additional components like clustering algorithms or further recommendation models and exploiting soft-computing techniques. We identified this suggestion in 33 studies.

3.9.2 Include more data or additional criteria

Another possibility is to improve the proposed approach by including more data or by increasing the number of criteria exploited by the recommendation algorithm. For example, it is possible to rely on external ontologies, contextual and content information, trust-related scores, and also consider additional criteria extracted from user reviews. This category of future works was mentioned in 24 studies.

3.9.3 Improve the evaluation

Some studies mention the fact that the evaluation performed by their authors was not enough complete or detailed because of different kinds of constrains. For this reason, it would be advisable to increase its trustfulness by executing it again considering more datasets, techniques, and evaluation metrics. In total, this problem was reported by 14 studies.

3.9.4 Identify significant criteria

A common issue associated with multicriteria recommenders is the noise introduced by redundant criteria, as discussed in Section 3.2.4. For this reason, some authors considered the identification of the most significant criteria as a future work. This suggestion was discussed in 13 studies.

3.9.5 Increase the scalability

A typical issue of multicriteria recommender systems is their limited scalability. This problem was highlighted by different studies in Section 3.2.6. Some researchers suggested to study how to increase it, for example by means of parallel computational paradigms. We identified this future work in 10 studies.

3.9.6 Consider different domains

Finally, 7 studies mentioned the necessity of validating the proposed approach in different domains, like it is usually done with traditional recommenders. However, this objective is difficult to achieve because of the limited availability of multicriteria datasets.

3.9.7 Other future works

Further categories of future works, mentioned in less than 5 studies each, include designing solutions for addressing the cold-start problem, as described in Section 3.2.5, performing experiments with adaptive recommenders, solving the issues related to preference elicitation, creating algorithms for explaining the recommendations, and performing an analysis of the related ethical problems.

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, and highlighting possible threats to validity.

4.1 Included studies

As reported in Figure 2, the earliest study included in this review, which is [P76], dates back to the year 2003. However, the field of multicriteria recommender systems started to be relatively widespread only from the year 2007, when influential studies like [P53] were published. We can also observe an increasing amount of publications, suggesting that this research topic is still popular, as recommender systems in general. In particular, the last few years were characterized by a higher number of studies related to specialized applications of multicriteria approaches, for example in tourism, health and care, and distance learning.

Regarding the quality of the included studies, summarized in Figure 3, we can highlight the fact that higher scores were assigned to works published in journals

with respect to conferences. This result is consistent with the conclusions of other systematic literature reviews conducted in related fields, like hybrid and linked data-based recommender systems [24, 7].

4.2 Research problems

By looking at the problems listed in Section 3.2, it is possible to observe that the most frequent one faced by researchers is data sparsity. This is a general issue of collaborative filtering recommender systems that causes a lower recommendation quality due to an insufficient amount of input data. However, it is reasonable to suppose that data sparsity is more severe in the context of multicriteria recommenders, because of the higher amount of expected ratings from each user. Possible solutions include the use of external information or the construction of latent models. In contrast, it is not clear what is the effect of using different rating elicitation methods.

Another typical research problem is discovering what are the optimal weights for each criterion. They may be computed globally or for each user with the objective of maximizing the recommendation accuracy. A second approach is to obtain the weights directly from the user, for example with decision making techniques.

A different but related issue is represented by data noise, caused by redundant criteria or dishonest ratings. Of course, in order to minimize the user's effort and the probability of obtaining inaccurate ratings, it is necessary to limit the number of criteria. However, identifying the most appropriate ones for a given domain is a difficult task even for an expert. In our opinion, this issue should be better investigated by future studies.

Accuracy is a characteristic required for any machine learning technique and, in the context of recommender systems, it is related to user satisfaction. Multicriteria recommenders can increase the personalization of the suggestions if they are capable of correctly identifying what are the most important criteria for each user. However, when a user is new to the system, the cold-start problem arises. This is a general issue of collaborative filtering recommenders and it is typically solved by creating hybrid solutions that consider content-based information.

Finally, the usage of multiple criterion results in algorithms that are less scalable. This problem can be addressed with clustering and dimensionality reduction techniques, as well as by exploiting distributed architectures.

4.3 Recommendation approaches

The vast majority of multicriteria recommender systems can be classified as collaborative filtering approaches, as reported in Table 6. For example, in heuristic-based methods, the ratings provided for each dimension are exploited together using a multidimensional distance metric, extending the traditional neighborhood-based recommendation technique. More complex heuristics rely on the aggregation of different similarities computed per criterion, possibly using weights specific for each user. Also in the context of model-based collaborative filtering, methods like matrix factorization are applied to each dimension and, then, their results are aggregated in a global predicted rating.

For this reason, almost all recommender systems included in this review can be considered hybrid, as they combine multiple collaborative filtering models, one for each criterion. Furthermore, such techniques are sometimes exploited together with content-based, knowledge-based, community-based, and demographic approaches, resulting in other forms of hybrid recommenders [5].

However, a limited number of studies included in this review deals with recommender systems that can be classified only as knowledge-based. Such systems exploit domain specific knowledge to match the available items with the user preferences. They were selected because we identified them as a form of multicriteria recommendation, even if the distinction between knowledge-based recommenders and traditional information retrieval methods is not always clear.

4.4 Multicriteria techniques

In Section 3.4, we reported that the most common technique exploited by multicriteria recommenders is the k -NN algorithm. This result is consistent with the recommendation approaches identified in the reviewed studies, as neighborhood-based methods represent a popular approach of collaborative filtering. k -NN may be enough to build a multicriteria recommender: the similarities among users or items can be directly computed with a multidimensional distance metric and they can be aggregated using a trivial approach, like the averaging function, or estimated by means of more complex techniques.

In contrast, multicriteria matrix manipulation methods represent a less common approach to collaborative filtering and they are often exploited to estimate unknown ratings for each criterion or to reduce the dimensionality of the problem. Again, matrix manipulation methods are usually combined with other methods designed to estimate importance weights for each criterion.

Different families of such techniques have been identified in the reviewed studies namely, regression analysis, neural networks, and genetic algorithms. It is possible to produce an overall rating using regression analysis if the other methods are not capable of dealing with multiple ratings on their own. A more recent alternative to regression analysis is represented by neural networks, that are exploited for learning the relative importance of criteria. Finally, genetic algorithms may be used to discover the weights of each criterion, but also to estimate good parameters for the recommendation model.

For reducing the complexity of the problem, some studies considered clustering algorithms in order to create cohesive groups of similar users or items. Such algorithms are usually applied as a first step, before other techniques capable of creating suggestions suitable for each cluster. In contrast, fuzzy logic may be exploited together with all the aforementioned methods for formally encoding the fact that ratings usually represent uncertain values.

The second most popular recommendation method after k -NN consists in decision analysis techniques. They are typically used to generate an ordered list of suggested items considering potential conflicts in user requirements. Decision analysis algorithms are often combined with k -NN, matrix manipulation, and clustering methods. However, some studies do not mention a specific recommendation technique to be exploited jointly with a decision analysis method because the

Technique	Benefit	Issue
k -NN	It can compute directly the predictions using a multidimensional metric.	If it is combined with other methods its scalability is reduced.
Decision analysis	It can analyze the importance that each user gives to the available criteria.	It may require additional data from the users apart their multicriteria ratings.
Fuzzy logic	It supposes that rating values may have a different meaning depending on the user.	It must be combined with additional techniques to actually recommend items.
Regression analysis	It is the most popular approach for discovering the individual weight of each criteria.	It must be combined with k -NN or matrix manipulation.
Clustering	It can create groups of users with similar rating behaviours and reduce the dimension of the problem inputs.	It must be exploited together with additional techniques.
Matrix manipulation	It can compute unknown ratings for each criteria in an efficient and effective way.	It usually require a final step to calculate an overall predicted rating.
Neural network	Recent approaches can directly predict the final ratings with an interesting accuracy.	It is often exploited to compute the weights of each criteria, adding more complexity.
Generic algorithm	It can automatically determine the optimal weight for each criteria and user.	It must be combined with additional techniques to predict the final ratings.

Table 11: The main benefits and issues of the reviewed techniques.

proposed approach is presented as a general framework and, thus, any recommendation method is suitable.

In Table 11, we report the main benefits and issues of the reviewed multicriteria methods. They could be considered as different compromises between the correctness of the suggestions and the complexity of the approach. Thus, it is of paramount importance being able to decide if, for a given task, it is more appropriate to foster the scalability of the system or the accuracy of the results.

Finally, in Figure 8, we summarize how such techniques could be exploited for constructing an ideal multicriteria recommender. In the preprocessing phase, common approaches designed to reduce the dimension of the input ratings are clustering and genetic algorithms. Also fuzzy logic may be applied at this point to transform the ratings in fuzzy numbers. During the rating prediction phase, it is possible to rely on k -NN, matrix manipulation methods, neural networks, and statistical models. These algorithms may already include a dimensionality reduction step and a way of obtaining an ordered list of suggestions. Finally, in the ranking phase, popular approaches are decision analysis, regression analysis, neural networks, and genetic algorithms. Please note that the first and third phase are, in general, optional and they add several layers of complexity to the developed system. Furthermore, some techniques may already perform these tasks during the rating prediction phase.

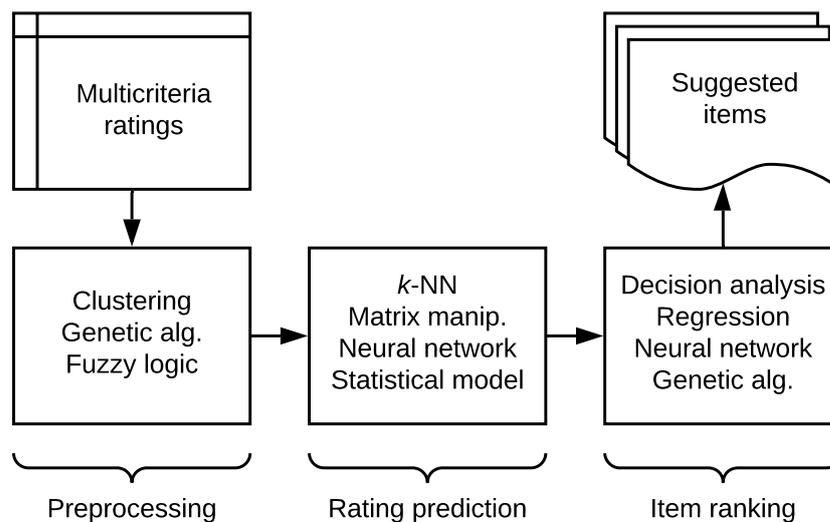


Fig. 8: This diagram represents the possible techniques exploited by an ideal multicriteria recommender, subdivided according to the recommendation phase. Content-based features and other less common approaches are not considered.

4.5 Application domains

As already observed in Section 3.5, multicriteria recommenders are frequently exploited in the tourism and travel domain. This result is expected, as many travel websites, like TripAdvisor⁷ and Bcn Restaurantes,⁸ rely on multicriteria ratings for suggesting items to their users. Even for someone that is not a domain expert, it is possible to identify some useful criteria that should be considered when evaluating a hotel or a restaurant.

For this reason, we can suppose that items belonging to complex domains can be more easily analyzed considering different dimensions and, therefore, they are better suited for multicriteria recommender systems. Another popular example is represented by consumer products, which can be evaluated considering different criteria based on their features, price, and quality.

On the other hand, domains like music or books are usually not exploited for creating a multicriteria recommender, probably because assigning detailed evaluations of such items is too difficult for someone that is not a domain expert.

Another factor that must be considered before performing further analysis is the availability of a certain multicriteria rating dataset in a particular domain. For example, many researchers used to collect ratings from the Yahoo! Movies website, but nowadays such service is no longer available.

The majority of the recommender systems presented in the reviewed studies are domain-independent, as they are only based on multicriteria ratings. However,

⁷ <https://www.tripadvisor.com>

⁸ <https://www.bcnrestaurantes.com>

some solutions designed for specific domains, like medical treatments and research papers, cannot be easily adapted to other scenarios.

4.6 Evaluation protocols

In general, performing the evaluation of a recommender system is a complex task, because the slightest change in the experimental protocol may produce radically different results [21]. It is not possible to directly compare the outcomes of different works because of the great variety of datasets, evaluation protocols, and metrics mentioned in the reviewed studies.

The multicriteria recommender systems were often evaluated using offline comparisons, as summarized in Table 8. This result is expected, as such an approach is less expensive than performing a user study, but it can be used to obtain useful information. However, in order to report more conclusive results, some authors decided to perform a user study. We noticed that it is difficult to involve a significant number of experimental subjects. In fact, only in 3 studies the number of participants was higher than a hundred people.

4.7 Evaluation metrics

The majority of the reviewed studies validated their approach using rating accuracy metrics, as reported in Table 9. This result is probably related to the fact that recommender systems were initially considered as tools designed to predict unknown ratings as accurately as possible. This evaluation approach produces results that are interesting from an academic point of view, but they may have a limited practical meaning.

In fact, as suggested in [12], it is not possible to display in an application all the items that are associated with the highest predicted ratings. Therefore, it is necessary to be capable of correctly creating lists of ranked items that will be shown to users. This aspect is considered by different studies, in which usage and ranking accuracy metrics are reported.

Other categories of metrics identified during the review include coverage and scalability, which were evaluated in a relatively low number of studies. It is interesting to notice that more recent evaluation dimensions, like diversity, novelty and serendipity, have not been considered, suggesting that their adoption in the recommender system community is still limited.

Finally, we also observed that the differences in the evaluation protocols followed by the reviewed studies do not allow a direct comparison of the results, even when widespread metrics, like MAE or precision, are exploited. For example, we identified a wide variability in the rating dataset, the splitting method between training and test set, and the definition of non-relevant items. This conclusion is consistent with the findings reported in [21].

4.8 Evaluation datasets

From the results presented in Section 3.8, we observe that different evaluation datasets have been created using data available in online portals that contain

multicriteria ratings. Such collections of ratings are usually not directly provided by the platforms, like the MovieLens datasets, but they are downloaded by researchers using web crawling methods. These approaches result in different versions of the same dataset that sometimes are not publicly available and they can only be obtained by contacting the original creators.

Consider, for example, Yahoo! Movies: this dataset was realized by adding multicriteria ratings to a collection of movie preferences initially provided by Yahoo! Research, quickly becoming one of the most widespread multicriteria gold standards. However, the Yahoo! Movies platform does not provide multicriteria ratings anymore, making impossible to recreate it. Similar observations are also valid for other datasets, as platforms may change over the years and original datasets may become unavailable. Anyway, we were able to locate two TripAdvisor dataset versions that are still publicly available on the Web.

It is interesting to notice that the authors of different studies decided to create new datasets, relying on real users and simulation tools, or to add additional criteria to existing datasets. This fact suggests that there is the need of realizing and publishing further multicriteria datasets.

4.9 Future works

We described the main future work directions reported by the authors of the reviewed studies in Section 3.9, in order to identify how it is possible to advance the multicriteria recommender field. The most common suggestion is to add to the proposed technique an additional component for improving its personalization. The authors of such recommenders usually observe that their methods could benefit from further preprocessing phases or additional components to be combined in a hybrid approach. Another possibility is to consider soft-computing techniques for expressing the uncertainty and vagueness of user ratings.

It is interesting to notice that some studies propose to improve multicriteria recommenders by considering further data, while other authors are more interested in reducing the number of available criteria. These suggestions are not in contrast, as they may be related to the exploited dataset. Therefore, correctly defining the rating criteria and better studying how such ratings are collected are of paramount importance for obtaining high quality recommendations.

Different works propose to also consider friendship relations among users, external knowledge bases and ontologies, and more contextual information. Social Recommender Systems (SRS) [11], Semantics-Aware Recommender Systems (SARS) [8], and Context-Aware Recommender Systems (CARS) [3] are relatively new recommendation approaches that could be exploited jointly with multicriteria techniques in order to create a more personalized experience. However, such additional data need to be managed in a proper way: for this reason, some researchers propose to also study the scalability of the available solutions.

As already discussed, recommender system evaluation is not an easy task. Therefore, different authors declare that they plan to conduct a more extensive experimentation of the proposed approach in future works. This frequent situation may be caused by the difficulties in obtaining multicriteria datasets or by the elevated costs necessary for involving real users in the evaluation process. A related

future work is to evaluate the same approach in different domains, as also this suggestion requires further datasets and experiments.

4.10 Threats to validity

In this section, we point out the main problems that we addressed while conducting this systematic literature review. We noticed that there is no strong agreement on a shared definition of multicriteria recommender systems, as there are some subtle differences but also some similarities with multiattribute content-filtering techniques [1]. For this reason, we only considered as multicriteria the recommender systems based on multicriteria ratings. We also included in our selection few studies describing methods that cannot be considered collaborative filtering approaches, but that exploit multicriteria ratings in a non-conventional way. Furthermore, a limited number of studies were initially selected by analyzing their title and abstract. However, we were later forced to exclude them even if they were potentially relevant because we were not able to retrieve their full-text version. We also noticed that some studies were described in different research papers, and, therefore, we analyzed these situations with attention for deciding which version to include. Sometimes we selected multiple works because we discovered that these contributions were only partially overlapped. Because of time constraints, many studies were analyzed by the first author, while other authors cross-checked a random sample of the results. We handled unclear cases with discussions involving all authors.

5 Conclusions

In the context of this systematic literature review, we analyzed a total number of 93 studies related to the topic of multicriteria recommender systems. We provided an answer to five different research questions, in order to identify which are the main problems that multicriteria recommenders aim to resolve, what recommendation approaches they usually exploit, and what are the most common machine learning and data mining techniques considered for realizing them. Furthermore, we investigated in which domains multicriteria recommenders can be applied, how they are evaluated in terms of experimental protocols, metrics, and rating datasets. Finally, we reported the most common suggestions available in selected studies for conducting future research works.

We identified data sparsity as the most frequent problem that researchers try to address. This issue could be caused by the fact that users are less willing to provide multicriteria ratings because they find this task difficult and time consuming, as they need to consider different aspects of the same item separately. A related research problem is understating what are the most appropriate criteria that should be considered in a particular domain. An excessive number of criteria will most likely result in a partial overlap of their meanings, thus causing a reduction of the prediction accuracy because of the additional noise in the input data. Further studies should better quantify what is the impact of this problem and how recommendation algorithms could reduce it.

We observed that almost all multicriteria recommenders rely on collaborative filtering techniques. This approach is, in fact, also popular in the context of single-criteria recommenders. A large portion of the reviewed studies propose hybrid recommendation methodologies, as they usually combine multiple runs of collaborative filtering algorithms, one for each criterion. However, we also observed a more recent trend of also relying on additional techniques, like content-based, knowledge-based, and community-based approaches. The motivating idea of such recommender systems is to increase the accuracy of the suggested items by exploiting further data behind pure multicriteria ratings.

Furthermore, we classified the machine learning and data mining techniques mentioned in the reviewed studies according to the recommendation phase in which they are considered. We observed that during the preprocessing phase different authors choose to reduce the dimension of the problem inputs by means of clustering and genetic algorithms. In the rating prediction phase, various collaborative filtering methods are considered, like k -NN, matrix manipulation techniques, neural networks, and statistical models. Finally, during the ranking phase, it is possible to rely on decision analysis, regression analysis, neural networks, and genetic algorithms. Such methods are combined in various ways, and sometimes integrated with the results of additional recommendation approaches.

We also investigated the domains that are usually considered for implementing a multicriteria recommender system. In the reviewed studies we frequently observed items related to the tourism and travel domain, for example hotels and restaurants. Another popular domain is represented by movies, probably due to the past availability of the Yahoo! Movies datasets. In contrast, domains that are often exploited in single-criteria recommenders, like music or books, are almost never considered for multicriteria approaches. Possible explanations of this finding could be the lack of proper evaluation datasets or the difficulties in defining meaningful criteria and obtaining the associated ratings.

From the analyzed experiments we observed that it is not possible to directly compare the results obtained with different recommendation approaches due to the extreme variability in the experimental protocols followed by their authors. Another critical point that emerged from this work is the lack of publicly available multicriteria rating datasets. Even if this problem could be partially mitigated with synthetic and crawled datasets, additional efforts should be done in order to create proper benchmarks in different domains and, thus, enabling a better reproducibility of the experimental results.

Finally, our findings indicate that future works in this field should better explore how to increase the utility of multicriteria recommenders by integrating community-based, knowledge-based, and context-aware approaches, according to the peculiar characteristics of the recommended items. For example, a restaurant recommender system would greatly benefit from context-aware information, a movie recommender could also leverage on semantic data, while a consumer product recommender may assign more importance to the opinions of trusted users. Another interesting point that should be better addressed in future works is defining what are the most appropriate criteria for each domain and how the ratings associated to that criteria should be collected in order to minimize user efforts, also considering soft-computing and linguistic techniques. Furthermore, it would be useful to explore novel algorithms for dynamically understanding the im-

portance that each user implicitly assigns to a criterion, for reducing the fatigue associated with the preference elicitation phase.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Adomavicius, G., Kwon, Y.: Multi-criteria recommender systems. In: *Recommender Systems Handbook*, pp. 847–880. Springer (2015). DOI 10.1007/978-1-4899-7637-6_25
2. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering* **17**(6), 734–749 (2005). DOI 10.1109/tkde.2005.99
3. Adomavicius, G., Tuzhilin, A.: Context-aware recommender systems. In: *Recommender Systems Handbook*, pp. 191–226. Springer (2015). DOI 10.1007/978-1-4899-7637-6_6
4. Balabanović, M., Shoham, Y.: Fab: content-based, collaborative recommendation. *Communications of the ACM* **40**(3), 66–72 (1997). DOI 10.1145/245108.245124
5. Burke, R.: Hybrid web recommender systems. In: *The Adaptive Web*, pp. 377–408. Springer Berlin Heidelberg (2007). DOI 10.1007/978-3-540-72079-9_12
6. Cruzes, D.S., Dyba, T.: Recommended steps for thematic synthesis in software engineering. In: *2011 International Symposium on Empirical Software Engineering and Measurement*. IEEE (2011). DOI 10.1109/esem.2011.36
7. Figueroa, C., Vagliano, I., Rocha, O.R., Morisio, M.: A systematic literature review of linked data-based recommender systems. *Concurrency and Computation: Practice and Experience* **27**(17), 4659–4684 (2015). DOI 10.1002/cpe.3449
8. de Gemmis, M., Lops, P., Musto, C., Narducci, F., Semeraro, G.: Semantics-aware content-based recommender systems. In: *Recommender Systems Handbook*, pp. 119–159. Springer (2015). DOI 10.1007/978-1-4899-7637-6_4
9. Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using collaborative filtering to weave an information tapestry. *Communications of the ACM* **35**(12), 61–70 (1992). DOI 10.1145/138859.138867
10. Gunawardana, A., Shani, G.: Evaluating recommender systems. In: *Recommender Systems Handbook*, 2 edn., chap. 8, pp. 265–308. Springer (2015). DOI 10.1007/978-1-4899-7637-6_8
11. Guy, I.: Social recommender systems. In: *Recommender Systems Handbook*, pp. 511–543. Springer (2015). DOI 10.1007/978-1-4899-7637-6_15
12. Herlocker, J.L., Konstan, J.A., Riedl, J.: Explaining collaborative filtering recommendations. In: *Proceedings of the 2000 ACM conference on Computer Supported Cooperative Work*. ACM (2000). DOI 10.1145/358916.358995

13. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems* **22**(1), 5–53 (2004). DOI 10.1145/963770.963772
14. Yi Hong, J., Ho Suh, E., Kim, S.J.: Context-aware systems: A literature review and classification. *Expert Systems with Applications* **36**(4), 8509–8522 (2009). DOI 10.1016/j.eswa.2008.10.071
15. Kitchenham, B., Charters, S.: Guidelines for performing systematic literature reviews in software engineering (2007)
16. Manouselis, N., Costopoulou, C.: Analysis and classification of multi-criteria recommender systems. *World Wide Web* **10**(4), 415–441 (2007). DOI 10.1007/s11280-007-0019-8
17. Park, D.H., Kim, H.K., Choi, I.Y., Kim, J.K.: A literature review and classification of recommender systems research. *Expert Systems with Applications* **39**(11), 10059–10072 (2012). DOI 10.1016/j.eswa.2012.02.038
18. Portugal, I., Alencar, P., Cowan, D.: The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications* **97**, 205–227 (2018). DOI 10.1016/j.eswa.2017.12.020
19. Quadrana, M., Cremonesi, P., Jannach, D.: Sequence-aware recommender systems. *ACM Computing Surveys* **51**(4), 1–36 (2018). DOI 10.1145/3190616
20. Ricci, F., Rokach, L., Shapira, B.: Recommender systems: Introduction and challenges. In: *Recommender Systems Handbook*, pp. 1–34. Springer (2015). DOI 10.1007/978-1-4899-7637-6_1
21. Said, A., Bellogín, A.: Comparative recommender system evaluation. In: *Proceedings of the 8th ACM Conference on Recommender Systems*, pp. 129–136. ACM (2014). DOI 10.1145/2645710.2645746
22. Wang, H., Lu, Y., Zhai, C.: Latent aspect rating analysis on review text data. In: *Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining*. ACM (2010). DOI 10.1145/1835804.1835903
23. Zhang, S., Yao, L., Sun, A., Tay, Y.: Deep learning based recommender system. *ACM Computing Surveys* **52**(1), 1–38 (2019). DOI 10.1145/3285029
24. Çano, E., Morisio, M.: Hybrid recommender systems: A systematic literature review. *Intelligent Data Analysis* **21**(6), 1487–1524 (2017). DOI 10.3233/IDA-163209

A Selected studies

Code	Author	Title	Year	Publication	Source
P1	Liu, L.; Mehandjiev, N.; Xu, D.-L.	Multi-criteria service recommendation based on user criteria preferences	2011	Fifth ACM Conference on Recommender Systems	ACM Digital Library
P2	Shambour, Q.; Lu, J.	A hybrid multi-criteria semantic-enhanced collaborative filtering approach for personalized recommendations	2011	International Conferences on Web Intelligence and Intelligent Agent Technology	ACM Digital Library
P3	Jannach, D.; Karakaya, Z.; Gedikli, F.	Accuracy improvements for multi-criteria recommender systems	2012	13th ACM Conference on Electronic Commerce	ACM Digital Library
P4	Hdioud, F.; Frikh, B.; Ouhbi, B.	Multi-criteria recommender systems based on multi-attribute decision making	2013	International Conference on Information Integration and Web-based Applications & Services	ACM Digital Library
P5	Choudhary, P.; Kant, V.; Dwivedi, P.	A particle swarm optimization approach to multi criteria recommender system utilizing effective similarity measures	2017	9th International Conference on Machine Learning and Computing	ACM Digital Library
P6	Musto, C.; de Gemmis, M.; Semeraro, G.; Lops, P.	A multi-criteria recommender system exploiting aspect-based sentiment analysis of users' reviews	2017	Eleventh ACM Conference on Recommender Systems	ACM Digital Library
P7	Park, Y.	Recommending personalized tips on new courses for guiding course selection	2017	South-East Conference	ACM Digital Library
P8	Sreepada, R. S.; Patra, B. K.; Hemando, A.	Multi-criteria recommendations through preference learning	2017	Fourth ACM IKDD Conferences on Data Sciences	ACM Digital Library

Code	Author	Title	Year	Publication	Source
P9	Zheng, Y.	Criteria chains: A novel multi-criteria recommendation approach	2017	22nd International Conference on Intelligent User Interfaces	ACM Digital Library
P10	Tallapally, D.; Sreepada, R. S.; Patra, B. K.; Babu, K. S.	User preference learning in multi-criteria recommendations using stacked auto encoders	2018	12th ACM Conference on Recommender Systems	ACM Digital Library
P11	Zheng, Y.; Dave, T.; Mishra, N.; Kumar, H.	Fairness in reciprocal recommendations	2018	26th Conference on User Modeling, Adaptation and Personalization	ACM Digital Library
P12	Niknafs, A.; Charkari, N. M.; Niknafs, A. A.	PROMETHEE-based recommender system for multi-sort recommendations in on-line stores	2008	Third International Conference on Digital Information Management	IEEE Xplore
P13	Hwang, C.-S.; Kao, Y.-C.; Yu, P.	Integrating multiple linear regression and multicriteria collaborative filtering for better recommendation	2010	International Conference on Computational Aspects of Social Networks	IEEE Xplore
P14	Liu, L.; Lecue, F.; Mehandjiev, N.; Xu, L.	Using context similarity for service recommendation	2010	IEEE Fourth International Conference on Semantic Computing	IEEE Xplore
P15	Shambour, Q.; Lu, J.	A framework of hybrid recommendation system for government-to-business personalized e-services	2010	Seventh International Conference on Information Technology: New Generations	IEEE Xplore
P16	Zarrinkalam, F.; Kahani, M.	A multi-criteria hybrid citation recommendation system based on linked data	2012	2nd International eConference on Computer and Knowledge Engineering	IEEE Xplore
P17	Boulkrinat, S.; Hadjali, A.; Mokhtari, A.	Enhancing recommender systems prediction through qualitative preference relations	2013	11th International Symposium on Programming and Systems	IEEE Xplore

Code	Author	Title	Year	Publication	Source
P18	Samatthiyadikun, P.; Takasu, A.; Maneeroj, S.	Bayesian model for a multicriteria recommender system with support vector regression	2013	IEEE 14th International Conference on Information Reuse & Integration	IEEE Xplore
P19	Bokde, D. K.; Girase, S.; Mukhopadhyay, D.	An approach to a university recommendation by multi-criteria collaborative filtering and dimensionality reduction techniques	2015	IEEE International Symposium on Nanoelectronic and Information Systems	IEEE Xplore
P20	Sharma, Y.; Bhatt, J.; Magon, R.	A multi-criteria review-based hotel recommendation system	2015	IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing	IEEE Xplore
P21	Asawarangsee, T.; Maneeroj, S.	A novel aggregation technique for multi-criteria recommendation	2016	13th International Joint Conference on Computer Science and Software Engineering	IEEE Xplore
P22	Ashley-Dejo, E.; Ngwira, S. M.; Zuva, T.	A context-aware proactive recommender system for tourist	2016	International Conference on Advances in Computing and Communication Engineering	IEEE Xplore
P23	Hassan, M.; Hamada, M.	Enhancing learning objects recommendation using multi-criteria recommender systems	2016	IEEE International Conference on Teaching, Assessment, and Learning for Engineering	IEEE Xplore
P24	Wijayanto, A.; Winarko, E.	Implementation of multi-criteria collaborative filtering on cluster using Apache Spark	2016	2nd International Conference on Science and Technology-Computer	IEEE Xplore

Code	Author	Title	Year	Publication	Source
P25	Hamada, M.; Odu, N. B.; Hassan, M.	A fuzzy-based approach for modelling preferences of users in multi-criteria recommender systems	2018	IEEE 12th International Symposium on Embedded Multicore/Many-core Systems-on-Chip	IEEE Xplore
P26	Mohamed, H.; Abdulsalam, L.; Mohammed, H.	Adaptive genetic algorithm for improving prediction accuracy of a multi-criteria recommender system	2018	IEEE 12th International Symposium on Embedded Multicore/Many-core Systems-on-Chip	IEEE Xplore
P27	Turk, A. M.; Bilge, A.	A robust multi-criteria collaborative filtering algorithm	2018	Innovations in Intelligent Systems and Applications	IEEE Xplore
P28	Manouselis, N.; Costopoulou, C.	Experimental analysis of design choices in multiattribute utility collaborative filtering	2007	International Journal of Pattern Recognition and Artificial Intelligence	ISI Web of Knowledge
P29	Yin, Z.; Yueting, Z.; Jiangqin, W.; Liang, Z.	Applying probabilistic latent semantic analysis to multi-criteria recommender system	2009	AI Communications	ISI Web of Knowledge
P30	Palanivel, K.; Sivakumar, R.	A study on implicit feedback in multicriteria e-commerce recommender system	2010	Journal of Electronic Commerce Research	ISI Web of Knowledge
P31	Dixit, V. S.; Mehta, H.; Bedi, P.	A proposed framework for group-based multi-criteria recommendations	2014	Applied Artificial Intelligence	ISI Web of Knowledge
P32	Mikeli, A.; Apostolou, D.; Despotis, D.	A new recommendation technique for interval scaled multi-criteria rating systems incorporating intensity of preferences	2015	Intelligent Decision Technologies	ISI Web of Knowledge

Code	Author	Title	Year	Publication	Source
P33	Hamada, M.; Hassan, M.	Artificial neural networks and particle swarm optimization algorithms for preference prediction in multi-criteria recommender systems	2018	Informatics	ISI Web of Knowledge
P34	Chen, D.-N.; Hu, P. J.-H.; Kuo, Y.-R.; Liang, T.-P.	A web-based personalized recommendation system for mobile phone selection: Design, implementation, and evaluation	2010	Expert Systems with Applications	ScienceDirect
P35	Huang, S.-l.	Designing utility-based recommender systems for e-commerce: Evaluation of preference-elicitation methods	2011	Electronic Commerce Research and Applications	ScienceDirect
P36	Liu, H.; He, J.; Wang, T.; Song, W.; Du, X.	Combining user preferences and user opinions for accurate recommendation	2013	Electronic Commerce Research and Applications	ScienceDirect
P37	Hu, Y.-C.	Nonadditive similarity-based single-layer perceptron for multi-criteria collaborative filtering	2014	Neurocomputing	ScienceDirect
P38	Li, Y.-M.; Chou, C.-L.; Lin, L.-F.	A social recommender mechanism for location-based group commerce	2014	Information Sciences	ScienceDirect
P39	Marin, L.; Moreno, A.; Isern, D.	Automatic preference learning on numeric and multi-valued categorical attributes	2014	Knowledge-Based Systems	ScienceDirect
P40	Nilashi, M.; bin Ibrahim, O.; Ithnin, N.	Hybrid recommendation approaches for multi-criteria collaborative filtering	2014	Expert Systems with Applications	ScienceDirect

Code	Author	Title	Year	Publication	Source
P41	Son, L. H.; Thong, N. T.	Intuitionistic fuzzy recommender systems: An effective tool for medical diagnosis	2015	Knowledge-Based Systems	ScienceDirect
P42	Ali, M.; Son, L. H.; Thanh, N. D.; Minh, N. V.	A neutrosophic recommender system for medical diagnosis based on algebraic neutrosophic measures	2017	Applied Soft Computing	ScienceDirect
P43	Amoretti, M.; Belli, L.; Zanichelli, F.	UTravel: Smart mobility with a novel user profiling and recommendation approach	2017	Pervasive and Mobile Computing	ScienceDirect
P44	Choudhary, P.; Kant, V.; Dwivedi, P.	Handling natural noise in multi criteria recommender system utilizing effective similarity measure and particle swarm optimization	2017	Procedia Computer Science	ScienceDirect
P45	Kermany, N. R.; Alizadeh, S. H.	A hybrid multi-criteria recommender system using ontology and neuro-fuzzy techniques	2017	Electronic Commerce Research and Applications	ScienceDirect
P46	Yuen, K. K. F.	The fuzzy cognitive pairwise comparisons for ranking and grade clustering to build a recommender system: An application of smartphone recommendation	2017	Engineering Applications of Artificial Intelligence	ScienceDirect
P47	Akçayol, M. A.; Utku, A.; Aydoğan, E.; Mutlu, B.	A weighted multi-attribute-based recommender system using extended user behavior analysis	2018	Electronic Commerce Research and Applications	ScienceDirect

Code	Author	Title	Year	Publication	Source
P48	Castillo, A.; Meer, D. V.; Castellanos, A.	ExUP recommendations: Inferring user's product metadata preferences from single-criterion rating systems	2018	Decision Support Systems	ScienceDirect
P49	Nilashi, M.; Ibrahim, O.; Yadegaridehkordi, E.; Samad, S.; Akbari, E.; Alizadeh, A.	Travelers decision making using online review in social network sites: A case on TripAdvisor	2018	Journal of Computational Science	ScienceDirect
P50	Núñez-Valdez, E. R.; Quintana, D.; Crespo, R. G.; Isasi, P.; Herrera-Viedma, E.	A recommender system based on implicit feedback for selective dissemination of ebooks	2018	Information Sciences	ScienceDirect
P51	Song, W.; Sakao, T.	An environmentally conscious PSS recommendation method based on users' vague ratings: A rough multi-criteria approach	2018	Journal of Cleaner Production	ScienceDirect
P52	Wasid, M.; Ali, R.	An improved recommender system based on multi-criteria clustering approach	2018	Procedia Computer Science	ScienceDirect
P53	Adomavicius, G.; Kwon, Y.	New recommendation techniques for multicriteria rating systems	2007	IEEE Intelligent Systems	Scopus
P54	Lee, H.-H.; Teng, W.-G.	Incorporating multi-criteria ratings in recommendation systems	2007	IEEE International Conference on Information Reuse and Integration	Scopus
P55	Hwang, C.-S.	Genetic algorithms for feature weighting in multi-criteria recommender systems	2010	Journal of Convergence Information Technology	Scopus
P56	Lousame, F. P.; Sanchez, E.	Multicriteria predictors using aggregation functions based on item views	2010	10th International Conference on Intelligent Systems Design and Applications	Scopus

Code	Author	Title	Year	Publication	Source
P57	Tangphoklang, P.; Maneeroj, S.; Takasu, A.	Advanced representative and dynamic user profile based on MCDM for multi-criteria RS	2010	IADIS International Conference Information Systems	Scopus
P58	Akhtarzada, A.; Calude, C. S.; Hosking, J.	A multi-criteria metric algorithm for recommender systems	2011	Fundamenta Informaticae	Scopus
P59	Lakiotaki, K.; Matsatsinis, N. F.; Tsoukias, A.	Multicriteria user modeling in recommender systems	2011	IEEE Intelligent Systems	Scopus
P60	Palanivel, K.; Sivakumar, R.	A study on collaborative recommender system using fuzzy-multicriteria approaches	2011	International Journal of Business Information Systems	Scopus
P61	Shambour, Q.; Lu, J.	Integrating multi-criteria collaborative filtering and trust filtering for personalized recommender systems	2011	IEEE Symposium on Computational Intelligence in Multicriteria Decision-Making	Scopus
P62	Samatthiyadikun, P.; Takasu, A.; Maneeroj, S.	Multicriteria collaborative filtering by Bayesian model-based user profiling	2012	IEEE 13th International Conference on Information Reuse & Integration	Scopus
P63	Premchaiswadi, W.; Poompuang, P.	Hybrid profiling for hybrid multicriteria recommendation based on implicit multicriteria information	2013	Applied Artificial Intelligence	Scopus
P64	Agathokleous, M.; Tsapatsoulis, N.	Learning user models in multi-criteria recommender systems	2014	Engineering Applications of Neural Networks	Scopus
P65	Bilge, A.; Kaleli, C.	A multi-criteria item-based collaborative filtering framework	2014	11th International Joint Conference on Computer Science and Software Engineering	Scopus

Code	Author	Title	Year	Publication	Source
P66	Hdioud, F.; Frikh, B.; Oubbi, B.	Bootstrapping recommender systems based on a multi-criteria decision making approach	2014	International Conference on Next Generation Networks and Services	Scopus
P67	Hu, Y.-C.	A multicriteria collaborative filtering approach using the indifference relation and its application to initiator recommendation for group-buying	2014	Applied Artificial Intelligence	Scopus
P68	Manouselis, N.; Kyrgiazos, G.; Stoitsis, G.	Exploratory study of multi-criteria recommendation algorithms over technology enhanced learning datasets	2014	Journal of E-Learning and Knowledge Society	Scopus
P69	Pinandito, A.; Ananta, M. T.; Brata, K. C.; Fanani, L.	Alternatives weighting in analytic hierarchy process of mobile culinary recommendation system using fuzzy	2015	ARPN Journal of Engineering and Applied Sciences	Scopus
P70	Sheha, Y. S.; Mahadevan, G.	A novel approach to personalized recommender systems based on multi criteria ratings	2015	Research Journal of Applied Sciences, Engineering and Technology	Scopus
P71	Bankshinategh, B.; Spanakis, G.; Zaiane, O.; ElAtia, S.	A course recommender system based on graduating attributes	2017	9th International Conference on Computer Supported Education	Scopus
P72	Goswami, A.; Dwivedi, P.; Kant, V.	Trust-enhanced multi-criteria recommender system	2017	Advances in Intelligent Systems and Computing	Scopus
P73	Hassan, M.; Hamada, M.	A neural networks approach for improving the accuracy of multi-criteria recommender systems	2017	Applied Sciences	Scopus

Code	Author	Title	Year	Publication	Source
P74	Leal, F.; González-Vélez, H.; Malheiro, B.; Burguillo, J. C.	Profiling and rating prediction from multi-criteria crowd-sourced hotel ratings	2017	31st European Conference on Modelling and Simulation	Scopus
P75	Majumder, G. S.; Dwivedi, P.; Kant, V.	Matrix factorization and regression-based approach for multi-criteria recommender system	2017	Information and Communication Technology for Intelligent Systems	Scopus
P76	Karacapilidis, N.; Hatzieleftheriou, L.	Exploiting similarity measures in multi-criteria based recommendations	2003	E-Commerce and Web Technologies	Springer Link
P77	Manouselis, N.; Costopoulou, C.	Preliminary study of the expected performance of MAUT collaborative filtering algorithms	2008	The Open Knowledge Society	Springer Link
P78	Matsatsinis, N. F.; Manarolis, E. A.	New hybrid recommender approaches: An application to equity funds selection	2009	Algorithmic Decision Theory	Springer Link
P79	Naak, A.; Hage, H.; Aïmeur, E.	A multi-criteria collaborative filtering approach for research paper recommendation in Papyrus	2009	E-Technologies: Innovation in an Open World	Springer Link
P80	Bitonto, P. D.; Laterza, M.; Roselli, T.; Rossano, V.	Multi-criteria retrieval in cultural heritage recommendation systems	2010	Knowledge-Based and Intelligent Information and Engineering Systems	Springer Link
P81	Maneroj, S.; Samatthiyadikun, P.; Chalermponpong, W.; Panthuwadeethorn, S.; Takasu, A.	Ranked criteria profile for multi-criteria rating recommender	2012	Information Systems, Technology and Management	Springer Link

Code	Author	Title	Year	Publication	Source
P82	Fan, J.; Xu, L.	A robust multi-criteria recommendation approach with preference-based similarity and support vector machine	2013	Advances in Neural Networks	Springer Link
P83	Jannach, D.; Zanker, M.; Fuchs, M.	Leveraging multi-criteria customer feedback for satisfaction analysis and improved recommendations	2014	Information Technology & Tourism	Springer Link
P84	Nilashi, M.; Ibrahim, O. B.; Ithnin, N.; Zakaria, R.	A multi-criteria recommendation system using dimensionality reduction and neuro-fuzzy techniques	2014	Soft Computing	Springer Link
P85	Chen, T.; Chuang, Y. H.	Fuzzy and nonlinear programming approach for optimizing the performance of ubiquitous hotel recommendation	2015	Journal of Ambient Intelligence and Humanized Computing	Springer Link
P86	Li, S. T.; Pham, T. T.; Chuang, H. C.; Wang, Z.-W.	Does reliable information matter? Towards a trustworthy co-created recommendation model by mining unboxing reviews	2015	Information Systems and e-Business Management	Springer Link
P87	Parveen, R.; Kant, V.; Dwivedi, P.; Jaiswal, A. K.	Enhancing recommendation quality of a multi criterion recommender system using genetic algorithm	2015	Mining Intelligence and Knowledge Exploration	Springer Link
P88	Jhalani, T.; Kant, V.; Dwivedi, P.	A linear regression approach to multi-criteria recommender system	2016	Data Mining and Big Data	Springer Link

Code	Author	Title	Year	Publication	Source
P89	Kant, V.; Jhalani, T.; Dwivedi, P.	Enhanced multi-criteria recommender system based on fuzzy Bayesian approach	2017	Multimedia Tools and Applications	Springer Link
P90	Ko, H.-G.; Ko, I.-Y.; Lee, D.	Multi-criteria matrix localization and integration for personalized collaborative filtering in IoT environments	2017	Multimedia Tools and Applications	Springer Link
P91	Leal, F.; Malheiro, B.; González-Vélez, H.; Burguillos, J. C.	Trust-based modelling of multi-criteria crowdsourced data	2017	Data Science and Engineering	Springer Link
P92	Ding, Y.; Li, S.; Yu, W.	Multi-criteria recommendation schemes based on factorization machines	2018	Cluster Computing	Springer Link
P93	Ding, Y.; Li, S.; Yu, W.; Wang, J.; Liu, M.	A unified neural model for review-based rating prediction by leveraging multi-criteria ratings and review text	2018	Cluster Computing	Springer Link