A Narrative Review of Multi-Criteria Recommender Systems: From Perspectives of Multi-Criteria Decision Making

Yong Zheng[a,*], David Xuejun Wang[b]

[a]Department of Information Technology and Management, Illinois Institute of Technology, Chicago, 60616, Illinois, USA
[b]Morningstar, Inc., 22 W Washington St, 60602 Chicago, Illinois, USA

ARTICLE INFO

Keywords:
Recommender System
Multi-Criteria Recommender Systems
Multi-Criteria Decision Making
Deep Learning
Preference Aggregation

ABSTRACT

Traditional recommender systems primarily rely on overall ratings, which may fail to capture the fine-grained nuances of user preferences. Multi-Criteria Recommender Systems (MCRS) address this limitation by incorporating user feedback across multiple aspects of the items to provide more accurate and personalized suggestions. While the integration of advanced deep learning techniques has significantly improved MCRS performance, the research landscape remains fragmented, lacking a unified theoretical framework that connects modern neural network architectures with classical Multi-Criteria Decision Making (MCDM) theories. To address this gap, this paper presents a conceptually-driven narrative review that introduces a novel architectural taxonomy for MCRS. We systematically classify existing methodologies into two paradigms: Multi-Stage architectures and Single-Stage architectures. Multi-Stage MCRS explicitly integrates MCDM principles, such as multi-attribute utility functions or Pareto dominance, by decomposing the recommendation pipeline into multi-criteria rating prediction and preference aggregation. Conversely, Single-Stage MCRS employs holistic, end-to-end modeling paradigms, such as neighborhood-based methods, deep tensor factorization, or multiview graph neural networks, in order to capture complex and non-linear dimensional interactions simultaneously. Moreover, this review examines the transition from relying on explicit criteria elicitation to mining implicit preferences from user-generated reviews, a shift that significantly broadens the real-world applicability of MCRS. We also discuss ongoing challenges in the field, such as data sparsity, integration with contextual situations, and model explainability. By bridging classical decision science with modern deep learning, this review provides a clear roadmap and theoretical insights for future MCRS research.

1. Introduction

Traditional recommender systems (RSs) assist users in alleviating information overload by providing personalized suggestions tailored to their preferences. However, conventional RSs primarily rely on a single overall rating, which fails to capture the fine-grained nuances of user preferences [2]. To address this limitation, Multi-Criteria Recommender Systems (MCRS) have emerged, allowing users to express their preferences across various item attributes [2, 53, 6, 94].

In real-world applications, many platforms explicitly collect such multi-criteria ratings, which enables multi-criteria recommendations. For instance, on the hotel booking platform TripAdvisor.com, users evaluate accommodations not only through an overall score but also based on specific criteria such as room quality, cleanliness, location, and services [28]. Similarly, dining reservation platforms like OpenTable.com collect fine-grained feedback on food quality, service, ambience, and value [88], as shown by Figure 1. More importantly, even if explicit multi-criteria ratings are not directly collected, the application of MCRS is no longer restricted. Recent advancements in text mining and sentiment analysis enable the extraction of implicit multi-criteria preferences directly from user-generated textual reviews [73, 56, 6]. This paradigm shift significantly broadens the applicability of MCRS, demonstrating that multi-criteria recommendation algorithms can be deployed in any domain as long as user reviews are available.

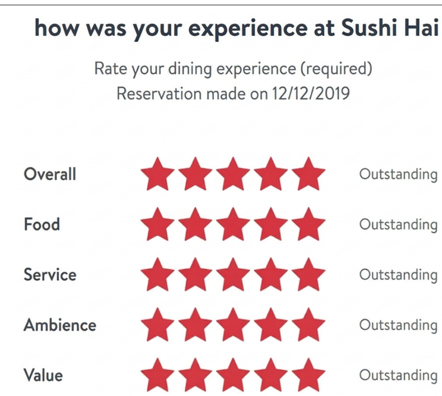

Figure 1: Example: Multi-criteria ratings on OpenTable

Over the past years, a variety of MCRS algorithms have been developed to leverage these multi-dimensional preferences. The most popular paradigm is the aggregation-based approach, which typically operates by first predicting the multi-criteria ratings and subsequently aggregating them

*Corresponding author
yzheng66@illinoistech.edu (Y. Zheng)
ORCID(s): 0000-0003-4990-4580 (Y. Zheng); 0000-0003-0724-7615 (D.X. Wang)

to estimate the overall rating [1, 93]. More recently, researchers have also introduced deep learning techniques into MCRS. By effectively capturing the complex and non-linear interactions between users and items, these advanced neural network architectures are utilized to build highly accurate recommendation models [57, 93]. This paper presents a conceptually-driven narrative review of MCRS, distinctively anchored in the classical principles of Multi-Criteria Decision Making (MCDM), aiming to provide a novel taxonomy and theoretical framework for the advances in MCRS.

1.1. The Missing Link: MCDM in MCRS

Fundamentally, the recommendation process in MCRS is aligned with the classical theories from MCDM [94]. Interestingly, the vast majority of existing MCRS algorithms belong to the category of *Multi-Stage MCRS*, where the recommendation pipeline is typically decomposed into predicting multi-criteria ratings and subsequently aggregating them into an overall score or ranking [94]. This aggregation stage heavily relies on the multi-attribute utility functions from MCDM methods [2, 21, 72].

Despite this evidence on MCDM in practical MCRS algorithms, the theoretical connection between the two fields remains fragmented. We argue that deeply understanding MCDM provides tremendous benefits for the further development of MCRS. For instance, classical MCDM principles can offer theoretically sound non-aggregation methods (e.g., Pareto dominance), inspire novel loss functions in neural networks, and enhance the explainability of the algorithms based on deep learning. Theoretically, almost all classical MCDM techniques could be reused to construct MCRS architectures; however, the vast majority of current literature remains heavily concentrated on aggregation-based utility models. By establishing a novel connection between classical MCDM theories and MCRS, our review provides a unified perspective on the relationship between MCDM and MCRS and identifies several promising research directions.

1.2. Limitations of Existing Surveys

In academic literature, review articles generally fall into three distinct methodological paradigms, each serving a different purpose [71]: (1) *Comprehensive or Systematic Literature Reviews (SLRs)* aim to exhaustively collect and quantitatively profile all available studies within a well-defined boundary, often adhering to strict and reproducible protocols; (2) *Scoping Surveys* attempt to rapidly map the broad concepts and identify the extent of research gaps in an emerging field [62]; and (3) *Narrative Reviews* are interpretive in nature. Rather than exhaustively enumerating papers, a narrative review selectively synthesizes milestone literature to construct novel theoretical frameworks, bridge disparate disciplines, and provide deep architectural insights [70].

In the field of MCRS, the vast majority of existing surveys predominantly adopt the comprehensive or scoping paradigms. For instance, Monti et al. [53] conducted a comprehensive PRISMA-based SLR to statistically profile the domain, categorizing 93 MCRS studies. Similarly, Al-Ghuribi and Noah [6] provided a scoping review focused on extracting multi-criteria preferences from textual reviews, while Rajput et al. [63] recently conducted an SLR on the integration of deep learning models in MCRS. Additionally, Pelissari et al. [61] presented a preliminary scoping review regarding the application of MCDM methods in general recommender systems, where actual data-driven MCRS only accounted for a small portion of their study. Instead, we revisited MCRS from the perspectives of MCDM and delivered an academic tutorial at the ACM IUI conference in 2023 [94].

While these comprehensive and scoping surveys provide valuable bibliometric statistics and effectively map the research landscape, they exhibit notable methodological limitations. Driven by the goal of exhaustive inclusion, they tend to mechanically categorize literature based on utilized algorithms, thereby lacking a high-level architectural synthesis. Regarding the intersection of decision science and recommendation technologies, although Pelissari et al. [61] initiated a preliminary statistical mapping, their scope was broadly oriented towards general RSs rather than dedicated MCRS. Their work mainly categorized the application of MCDM into basic functional phases (e.g., preference elicitation) and entirely omitted the integration, impact, and architectural complexities of modern neural network-based MCRS methodologies.

To fill this gap, this paper adopts a conceptually-driven narrative review approach. By stepping away from the mechanical constraints of comprehensive SLRs, this narrative approach enables us to systematically select representative literature and focus on the connections between MCDM and MCRS. It allows us to construct the proposed architectural taxonomy and highlight how classical decision science can offer theoretical scaffolding to interpret, guide, and inspire the design of future MCRS.

1.3. Contributions and Organization

By bridging decision science with modern recommendation techniques, the main contributions of this narrative review are three-fold:

- We establish a theoretical connection between modern MCRS architectures and classical MCDM paradigms, providing a unified perspective for understanding existing MCRS methodologies.
- We propose a novel pipeline-based architectural taxonomy that categorizes MCRS into Multi-Stage and Single-Stage paradigms. Unlike traditional classifications based primarily on underlying algorithms, the proposed taxonomy emphasizes computational workflows, architectural design choices, and the mechanisms through which multidimensional preferences are modeled and utilized.
- We synthesize recent advances in MCRS, including graph learning, contrastive learning, hyperbolic representation learning, causal modeling, and emerging

MCDM-aware architectures. Furthermore, we provide a critical discussion of the strengths, limitations, and applicable boundaries of different MCRS paradigms, and outline promising future research directions such as explainable MCRS, neural–MCDM integration, implicit preference extraction, and the fusion of MCRS with other recommendation paradigms.

The remainder of this paper is organized as follows. Section 2 defines the fundamentals and data modalities of MCRS. Section 3 provides a preliminary study of MCDM theories. Section 4 introduces our proposed architectural taxonomy. Section 5 and Section 6 thoroughly review the Multi-Stage and Single-Stage MCRS architectures, respectively. Section 7 discusses open challenges and future directions, followed by the conclusion in Section 8.

2. Fundamentals of MCRS

This section outlines the fundamentals of MCRS. First, we provide the formal definition of the MCRS problem and declare the scope of this review. Subsequently, we highlight the paradigm shift in modern MCRS methodologies, i.e., the transition from relying on explicitly elicited numerical ratings to extracting implicit multi-criteria preferences directly from user-generated textual reviews. Finally, we summarize the benchmark datasets that fuel these data-driven models, providing an essential data perspective for the algorithmic taxonomies discussed in the following sections. Moreover, we introduce the MCRecKit library which is an open-source library for MCRS.

2.1. MCRS: Definitions and Scopes

Formally, the recommendation problem in traditional RSs can be defined as a mapping function that predicts an overall rating based on the interactions between users and items:

$$R : Users \times Items \rightarrow R_0 \tag{1}$$

where R_0 denotes the overall rating representing how much a user likes an item. In contrast, MCRS extends this paradigm by additionally considering a user's preferences across multiple specific aspects or criteria of the items [2]. Assuming there are k distinct criteria, the mapping function for MCRS is defined as:

$$R : Users \times Items \rightarrow R_0 \times R_1 \times R_2 \times \cdots \times R_k \tag{2}$$

where R_1 to R_k represent the ratings given to each individual criterion. Table 1 presents an example of a multi-criteria rating matrix collected from the restaurant reservation platform OpenTable.com [88]. In this scenario, in addition to the overall rating, users express their fine-grained preferences regarding food quality, dining services, ambience of the space, and value of the money. As shown in the table, the following research problem in modern MCRS revolves around predicting the missing overall rating (or multi-criteria ratings) for a specific user-item pair (e.g., User U_3 on Item T_1) based on the historical rating data. Usually, the predicted overall ratings on the candidate items can be utilized to sort and rank candidates to produce the top-N recommendation list.

It is important to distinguish the scope of MCRS discussed in this review from two other prevalent research paradigms that occasionally use similar terminology but differ in their data foundations and methodologies. First of all, in classical decision science, literature sometimes utilizes the term "multi-criteria recommendation" or "recommender systems", such as [58, 77]. However, their core philosophy is focused on *decision analysis* rather than *preference prediction*. For instance, when recommending a tourism destination, classical MCDM methods typically evaluate alternatives based on objective criteria (e.g., cost, distance, climate) combined with explicitly elicited weights from the user [9, 22, 67]. This process relies entirely on the inherent attributes of the alternatives, without utilizing any historical user ratings. In addition, within the recommender systems community, some multi-attribute based recommendations were sometimes labeled with terms like "multi-criteria" and "recommendations", but they are out of the scope in our discussions, such as [10, 27]. While they utilize multiple attributes, these systems lack the users' preferences on these specific attributes, i.e., there are no multi-criteria ratings provided by users in the dataset.

By contrast, the MCRS methodologies reviewed in this paper are strictly defined as recommender systems utilizing multi-criteria preferences. Our scope is exclusively focused on machine learning systems that leverage historical multi-criteria rating matrices, as exemplified in Table 1, where users express their subjective and fine-grained preferences across multiple dimensions, to predict unknown ratings and subsequently generate Top-N recommendations.

2.2. Paradigm Shift: From Explicit Ratings to Implicit Reviews

In addition to traditional RSs, there are several specialized RSs, such as cross-domain RSs [43], context-aware RSs [3, 90, 51], group RSs [50, 18], multi-objective RSs [95]. Most of them may need additional inputs, which limits the scopes of their applications. For example, context-aware RSs need ratings in multiple contextual situations, where cross-domain RSs require algorithms to take user preferences from multiple domains into account.

Among them, MCRS emerged to capture fine-grained user preferences across multiple aspects of the items. Historically, however, the algorithmic development and real-world applicability of traditional MCRS have faced severe bottlenecks. Early MCRS architectures were strictly confined to scenarios where platforms provided explicit multi-criteria ratings. However, platforms that explicitly collect such multi-criteria ratings are limited, with TripAdvisor, OpenTable, and Yahoo! Movies being the popular examples. Therefore, the development of MCRS was heavily restricted by this data scarcity, limiting its broad applicability across diverse domains.

DOI:10.1016/j.neucom.2026.134673

Table 1
Example: Multi-Criteria Rating Matrix from OpenTable.com

User	Item	Overall Rating	Food	Service	Ambience	Value
U_1	T_3	4	4	3	4	4
U_2	T_2	3	3	3	3	3
U_3	T_1	?	?	?	?	?

To alleviate the dependency on explicit criterion ratings, MCRS research has gradually shifted toward extracting implicit multi-criteria preferences from unstructured user-generated content, driven by rapid advances in natural language processing (NLP). Early studies demonstrated that aspect-level ratings could be inferred directly from textual reviews through latent aspect rating analysis, topic modeling, and opinion mining techniques [73, 74, 32]. These approaches typically assume that user reviews contain hidden aspects corresponding to evaluation criteria and employ probabilistic topic models to jointly uncover latent aspects, aspect-specific sentiments, and user preferences. Thus, criterion-level ratings can be estimated even when users provide only overall ratings and textual comments.

Subsequent research further incorporated phrase-level sentiment analysis and aspect-based sentiment analysis (ABSA) to identify criterion-specific opinions and sentiments from reviews, enabling the construction of fine-grained preference representations [80, 56, 42]. Unlike latent topic models, ABSA-based approaches explicitly identify aspect terms (e.g., food, service, cleanliness, or value) and estimate the sentiment polarity associated with each aspect. This allows systems to directly map review content to predefined criteria and generate more interpretable criterion-level preference profiles.

More recently, advances in transformer-based models and large language models have substantially improved the extraction of aspect-level preferences, implicit criteria detection, preference completion, and criteria-level explanation generation from reviews and conversational interactions [36, 55, 78, 64]. Transformer architectures leverage contextual language representations to better capture semantic relationships between aspects and opinions, while large language models can perform instruction-guided preference extraction, infer missing criterion ratings, and generate structured user preference profiles from unstructured reviews and dialogues. These developments have significantly broadened the applicability of MCRS by enabling criterion-level preference modeling in domains where only overall ratings, textual reviews, comments, or conversational data are available.

Despite this progress, most existing studies focus on extracting, completing, or representing criterion-level information from unstructured data rather than directly developing MCRS algorithms. Moreover, challenges such as criterion-schema alignment, incomplete criterion coverage, and preference sparsity remain open research problems when transforming textual evidence into structured multi-criteria rating matrices. Therefore, while implicit preference extraction has become an important enabling technology for MCRS, a comprehensive review of NLP, ABSA, review mining, and LLM-based extraction techniques falls beyond the primary scope of this paper. Instead, this review focuses on MCRS methodologies and architectures after multi-criteria preference information has been obtained, whether through explicit ratings or implicit extraction mechanisms.

2.3. Open Resources

While the previous section highlighted the paradigm shift toward extracting implicit preferences from textual reviews, there is currently a lack of standardized, open-source datasets that provide ready-to-use, pre-extracted multi-criteria rating matrices from these reviews. To facilitate reproducible research and algorithmic benchmarking, the evaluation of MCRS predominantly relies on public datasets containing explicitly elicited multi-criteria ratings [53].

Obtaining such datasets has been challenging due to the limited number of platforms featuring explicit multi-criteria ratings. In recent years, however, researchers have published several valuable benchmark datasets via web crawling or user surveys. The most prominent open datasets utilized in MCRS research include:

- **Yahoo! Movies:** This is one of the earliest and most widely used benchmark datasets in the MCRS domain [53]. Originally collected by Jannach et al. [40], it contains explicitly elicited preferences on movies. Users evaluated movies based on four specific criteria: *direction, story, acting,* and *visual effects*, in addition to providing an overall rating. There are no public URLs to download the data set, but the datasets are available upon request from the authors.

- **TripAdvisor:** As a leading platform in the tourism and hospitality domain, TripAdvisor has served as the source for several MCRS datasets crawled by different researchers [53]. A highly utilized version, also provided by Jannach et al. [40], includes hotel evaluations across seven fine-grained criteria: *value for money, rooms, location, cleanliness, check-in/front desk, service,* and *business services*, alongside the overall rating. Data are available upon request from the authors.

- **OpenTable[1]:** To address the lack of high-quality, multi-criteria datasets in the dining recommendation

[1] https://www.kaggle.com/datasets/irecsys/opentable-data-with-multi-criteria-ratings

domain, Zheng [88] recently released the OpenTable dataset crawled from OpenTable.com. Diners evaluate restaurants based on four critical criteria: *food quality, service, ambience,* and *value*, making it an excellent modern benchmark for MCRS.

- **ITM-Rec[2]:** In the domain of project recommendations in education [83], Zheng introduced the ITM-Rec dataset [87]. It collects students' preferences regarding educational project topics or data based on three criteria: *app* (the appeal of the application domain), *data* (the ease of data preprocessing), and *ease* (the overall ease of the project). Moreover, ITM-Rec is significantly enriched with *contextual information* (e.g., academic semesters, course types, COVID-19 lockdown periods) and explicitly includes *group ratings and group compositions*, making it a uniquely versatile testbed for evaluating recommendations in fused scenarios, such as contextual multi-criteria recommendations [91, 23, 4], multi-criteria group recommendations [84, 47], and so forth.

Recent efforts have produced open-source libraries to support reproducible MCRS research. Traditional libraries like Surprise and LensKit focus on single-rating tasks and cannot handle multi-dimensional ratings. To address this, Zheng et al. introduced MCRecKit [98], the first Python-based, PyTorch-compatible library specifically for MCRS. MCRecKit includes automated data loaders for benchmark datasets (e.g., Yahoo!Movies, OpenTable, ITM-Rec) and a unified configuration environment for flexible multi-stage pipelines, enabling criteria-level training, aggregation, and GPU-accelerated multi-criteria ranking. Such toolkits effectively bridge the gap between theoretical MCRS design and practical large-scale implementation.

3. Preliminary: Multi-Criteria Decision Making

Before categorizing the modern algorithmic architectures of MCRS, it is essential to establish a solid theoretical foundation by reviewing the principles of classical MCDM. This section first introduces the fundamental definitions of MCDM, followed by an outline of its methodological taxonomy. Finally, we establish a conceptual mapping between classical decision science and multi-criteria recommender systems.

3.1. Introduction to MCDM

MCDM is a prominent subarea of operations research and management science, designed to assist decision-makers in evaluating and prioritizing alternatives when multiple, often conflicting, criteria must be considered simultaneously [37, 9, 67, 45]. Generally, the MCDM paradigm is categorized into two primary branches: Multi-Objective Decision-Making (MODM) and Multi-Attribute Decision-Making (MADM) [45]. MODM typically deals with continuous problems involving mathematical optimization and an infinite number of feasible solutions. In contrast, MADM is utilized for discrete problems featuring a finite, predefined set of alternatives. Since recommendation tasks inherently involve selecting from a list of candidate items (e.g., movies, hotels, or restaurants), the theoretical foundation of MCRS is rooted in the MADM paradigm.

According to classical decision science, despite the vast diversity of specific techniques, the ingredients of an MCDM problem are relatively straightforward. Formally, it requires at least one decision-maker, a finite set of available alternatives (or actions), and a family of evaluation criteria used to judge the performance of the alternatives [25, 29].

Furthermore, Bernard Roy, a pioneer in decision science, systematically categorized the ultimate goals of decision-aiding into distinct "decision problematics". When evaluating a set of alternatives based on multiple criteria, the MCDM process generally aims to solve one of the following core problematics [2, 14]. In the *choice problematic*, the goal is to select a single best alternative or a highly restricted subset of the most "satisfactory" alternatives from the entire candidate pool. By contrast, *sorting problematic* involves assigning each alternative into one of several pre-defined and preference-ordered categories. For example, classifying candidate alternatives into groups such as "accepted", "rejected", or "to be further considered" [54]. Finally, the *ranking problematic* aims to order the alternatives comprehensively from the most preferred to the least preferred.

These foundational problematics essentially dictate the mathematical output of an MCDM model. While classical MCDM methods can be applied to address all these problematics, the vast majority of existing MCRS research predominantly focuses on the ranking problematic (i.e., top-N recommendation task in RSs) [2]. Understanding this conceptual mapping lays the critical groundwork for comprehending how classical operations research algorithms can be adapted to fulfill modern recommendation tasks.

3.2. Methodological Taxonomy of MCDM

Over the past decades, operations research has produced hundreds of MCDM algorithms, each characterized by distinct mathematical axioms and decision philosophies [45]. To systematically navigate this vast landscape and provide theoretical support for modern recommender systems, we introduce the following popular taxonomy in MCDM. This taxonomy categorizes classical MCDM methodologies into four distinct families [25, 29]:

- **Value and Utility Theory Approaches:** Rooted in multi-attribute utility theory and multi-attribute value theory, this family operates on the principle of full compensation, where poor performance on one criterion can be compensated by high performance on another [29]. The core idea is to synthesize a decision-maker's multi-dimensional preferences into a single,

[2] https://www.kaggle.com/datasets/irecsys/itmrec

Table 2
Conceptual Mapping between MCDM and MCRS

MCDM Terminology	MCRS Terminology	Explanations
Decision-Maker	Active User	The user seeking decision support or recommendations.
Alternatives (or Actions)	Candidate Items	The discrete set of choices or candidates for evaluations.
Criteria	Aspects of items (Criteria)	The multiple evaluation aspects or perspectives.
Performance / Utility Values	Multi-Criteria Ratings	The explicit or implicit scores as user preferences.
Decision Process (Ranking)	Recommendation Generation	The process of the top-N recommendations.

global utility score for each alternative via an aggregation function (e.g., a weighted sum) [25]. Prominent methods in this category include the Analytic Hierarchy Process (AHP), Weighted Sum Model (WSM), and the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) [45].

- **Outranking Models:** Unlike utility-based methods, outranking approaches do not construct a single global score. Instead, they rely on pairwise comparisons to build a preference relation (an "outranking" relation) indicating whether one alternative is at least as good as another [25, 15]. These methods are typically non-compensatory or partially compensatory. They evaluate the strength of the assertion a outranks b using two core concepts: *concordance* (a majority of criteria must support the assertion) and *discordance* (no single criterion should strongly oppose or "veto" the assertion) [29]. Classic algorithms in this family include the ELECTRE and PROMETHEE series [45].
- **Preference Disaggregation Models:** Traditional MCDM methods typically require decision-makers to explicitly provide parameters (e.g., criteria weights or thresholds) *a priori*. In contrast, preference disaggregation is an *a posteriori* approach that essentially acts as reverse-engineering [39]. It infers the global preference model, such as marginal utility functions, indirectly from a set of holistic, global decision examples or historical rankings provided by the user [25, 39]. The most representative algorithm in this family is the UTA (Utilités Additives) method, which utilizes ordinal regression techniques to estimate these implicit preferences [38, 29].
- **Multi-Objective Optimization (MOO):** This mathematical programming approach treats each criterion as a separate objective function to be optimized simultaneously [19, 75, 95]. Because these objectives frequently conflict, there is rarely a single optimal solution. Instead, MOO aims to identify a set of *Pareto-optimal* (or non-dominated) solutions, representing the best possible trade-offs where no objective can be improved without degrading another [75, 95].

3.3. Mapping MCDM to MCRS

Having outlined the theoretical foundations and methodological taxonomy of classical MCDM, this preliminary study tries to establish a formal conceptual bridge between operations research and recommendation technologies. By drawing analogies between these two disciplines, we can translate classical MCDM frameworks into modern MCRS [61, 2].

More specifically, an item recommendation process in MCRS can be viewed as a multi-criteria decision-making process [6, 94]. To systematically demonstrate this synergy, Table 2 explicitly defines the formal mapping of terminologies between these two domains [2, 61, 6].

As detailed in Table 2, the individual seeking decision support corresponds directly to the active user in MCRS. The discrete set of feasible choices evaluated in MCDM are equivalent to the candidate items in RSs. Therefore, the multiple criteria utilized in MCDM map directly to the specific aspects or attributes of an item (e.g., *food, service, ambience* in dining recommendations) or criteria in MCRS. In this context, the performance score of an alternative on a specific criterion mathematically translates to the explicit or implicit criteria rating provided by a user to an item. Ultimately, the MCDM goal of aiding the decision-maker in making a choice aligns with the computational process of generating the top-N recommendation list.

Beyond the mapping of terminologies, it is essential to recognize that the methodological taxonomy introduced in Section 3.2 provides the theoretical scaffolding for almost all existing MCRS algorithms. For instance, the widely adopted "aggregation function" approach in MCRS could be linear regression [1], non-linear function like support vector regression [82], or complex multi-layer perceptrons [93, 57]. All these aggregation approaches are conceptually rooted in the family of multi-attribute utility theory, aiming to synthesize criteria ratings into a single overall rating. Furthermore, modern data-driven MCRS that learn latent weights or network parameters from historical multi-criteria rating matrices are essentially executing large-scale preference disaggregation. Recently, the concepts and techniques in MOO and outranking philosophies have also inspired novel recommendation frameworks, such as multi-criteria ranking based on the concept of Pareto dominance [93].

By establishing the mapping in this section, the connection between operations research and recommender systems becomes clearer. Equipped with this theoretical toolkit, we now proceed to systematically review and categorize the algorithmic architectures of modern MCRS in the subsequent sections.

DOI:10.1016/j.neucom.2026.134673

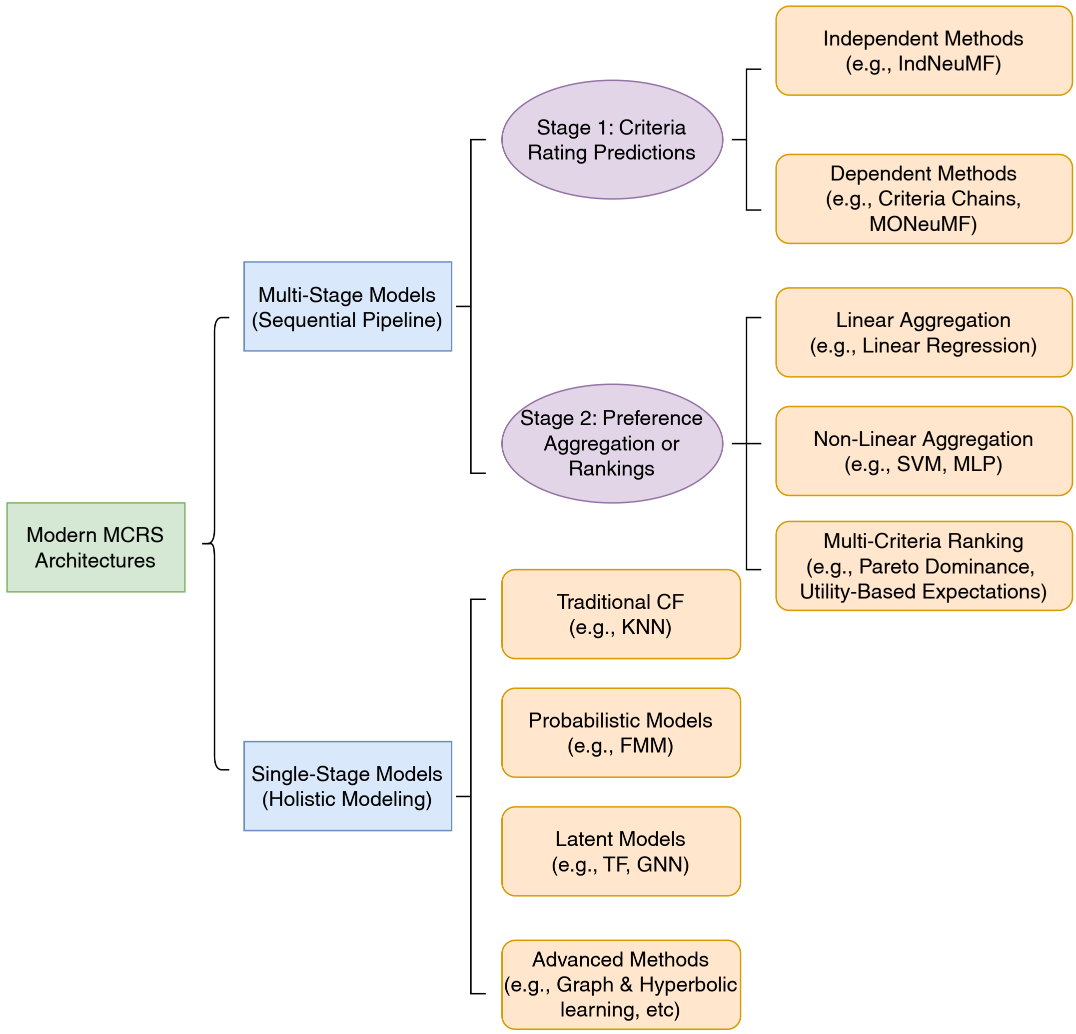

Figure 2: The Proposed Pipeline-Based Architectural Taxonomy for Modern MCRS

4. A Novel Taxonomy for MCRS

In 2010, Adomavicius et al. categorized MCRS algorithms based on the mathematical formation of their utility functions, dividing them into two traditional classes: *heuristic-based* (or memory-based) and *model-based* techniques [2]. In the traditional memory-based paradigm, these MCRS algorithms rely on the K-Nearest Neighbor (KNN) based collaborative filtering (CF), where systems typically compute user or item similarities on each criterion separately and aggregate them into a single similarity metric, or directly calculate multidimensional distances [2]. In other words, multi-criteria ratings were utilized to seek better neighborhood in CF. By contrast, model-based approaches, such as probabilistic modeling [66] or multilinear singular value decomposition [48], attempt to construct a predictive model to estimate unknown overall ratings from observed data.

While this conventional taxonomy was instrumental during the early evolution of MCRS, it exhibits notable limitations in the modern era of deep learning and complex neural architectures. Categorizing algorithms by their underlying mathematical techniques fails to capture the intricate data workflows and structural designs of modern systems. For instance, an advanced deep learning framework could simultaneously utilize both memory-like distance metrics and complex predictive models. The traditional heuristic versus model-based dichotomy has become increasingly insufficient for systematically profiling the state-of-the-art MCRS landscape.

To address this gap and provide a more structural perspective, we propose a novel, pipeline-based architectural taxonomy for MCRS, inspired by recent advancements and open-source MCRS frameworks in the field [98, 94]. Instead of focusing solely on the local utility functions, our taxonomy categorizes modern MCRS algorithms based on their computational pipelines and data routing mechanisms.

Table 3
Summary of the proposed architectural taxonomy of MCRS.

Paradigm	Subcategory	Section	Representative Methods
Multi-Stage MCRS	Independent criteria prediction	Section 5.1	KNN-based CF [1], MF [85], SVD [33], Autoencoder [11], IndNeuMF [93]
Multi-Stage MCRS	Dependent criteria prediction	Section 5.1	Criteria Chains [81], MONeuMF [57]
Multi-Stage MCRS	Aggregation-based recommendation	Section 5.2	Linear Regression [1], SVR [82], MLP [93, 57]
Multi-Stage MCRS	Multi-criteria ranking	Section 5.2	Pareto ranking [93, 96, 97], Utility-based models [85, 86]
Single-Stage MCRS	Neighborhood-based approaches	Section 6.1	KNN-based CF [1], TOPSIS/OWA-enhanced CF [8]
Single-Stage MCRS	Probabilistic models	Section 6.2	FMM [69, 66]
Single-Stage MCRS	Latent models	Section 6.3	MSVD [48], Contextual tensor factorization [35]
Single-Stage MCRS	Graph-based MCRS	Section 6.4.1	CPA-LGC [59], CA-GF [60], D-MGAC [26], TaTriGR [52]
Single-Stage MCRS	Hyperbolic and causal MCRS	Section 6.4.2	HMCR [30], MCCR [31]

Specifically, we classify the modern MCRS architectures into two paradigms:

- **Multi-Stage MCRS Architectures:** These models operate on a sequential pipeline. They explicitly divide the recommendation process into sequential phases, typically predicting intermediate multi-criteria ratings first, followed by a distinct preference aggregation or multi-criteria ranking stage [98, 94].
- **Single-Stage MCRS Architectures:** These models operate holistically. They directly leverage multi-criteria data to generate final recommendations or overall rating predictions in a unified, simultaneous process, without necessitating the explicit prediction of intermediate multi-criteria ratings [94].

Figure 2 illustrates this newly proposed architectural taxonomy, outlining the hierarchical sub-branches within each paradigm. To improve the readability of the proposed taxonomy, Table 3 summarizes the relationships among the architectural paradigms, their corresponding sections, representative algorithms, and key references discussed throughout this review. This mapping provides readers with a concise overview of the proposed taxonomy and facilitates navigation across different categories of MCRS methods. By shifting the classification standard from mathematical formulations to architectural pipelines, this taxonomy integrates classical MCDM theories with modern neural network designs. In the following sections, we will systematically unpack this taxonomy, discussing detailed algorithmic workflows of Multi-Stage architectures (Section 5) and Single-Stage architectures (Section 6), respectively.

5. Multi-Stage MCRS: Sequential Pipeline

As illustrated in our proposed taxonomy shown by Figure 2, multi-stage MCRS architectures explicitly decompose the recommendation task into a sequential pipeline. This paradigm aligns with classical MCDM processes, where the evaluation of alternatives across multiple criteria precedes the final decision synthesis. The vast majority of MCRS literature falls into this category, operating through two distinct phases: *Criteria Rating Prediction* (Stage 1) and *Preference Aggregation or Ranking* (Stage 2).

5.1. Stage 1: Criteria Rating Prediction

The primary objective of the first stage is to estimate the missing multi-criteria ratings for unobserved user-item pairs. Take the example shown in Table 1, the system tries to predict U_3's multi-criteria ratings on the item T_1 first and then aggregates these predicted ratings afterwards. Depending on whether the latent dependency among different criteria is considered, predictive models in this stage generally diverge into two distinct methodologies.

Independent Predictions. Early multi-stage systems typically assume that user preferences on different criteria are mutually independent. Accordingly, they deploy separate traditional collaborative filtering models (e.g., matrix factorization [85], SVD [33], KNN based CF [1, 34]) for each criterion to predict the corresponding ratings independently. In the deep learning era, this approach has been naturally extended to independent neural network models. For example, Batmaz et al. proposed to utilize autoencoder models to predict multi-criteria ratings separately [11]. Zheng et al. used neural matrix factorization (NeuMF) to estimate multi-criteria ratings independently, where this approach was named as independent NeuMF (IndNeuMF) [93], and the model architecture can be visualized by Figure 3a. We

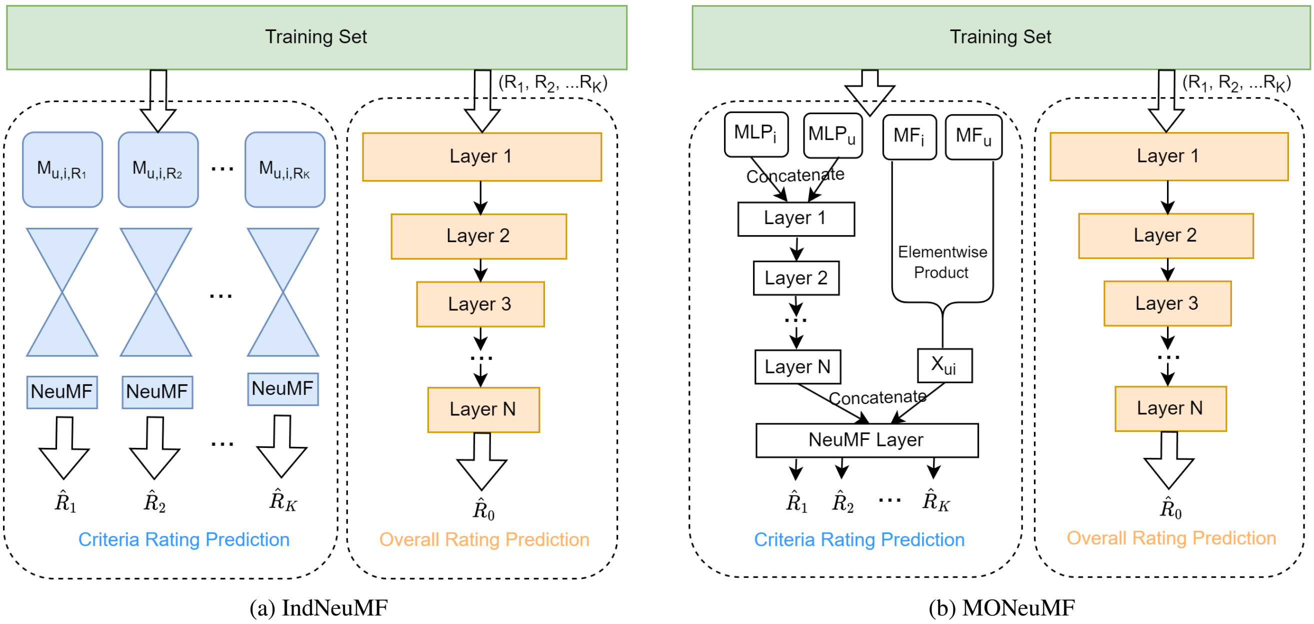

Figure 3: The architectures of deep multi-criteria recommendation models: (a) IndNeuMF, which utilizes independent NeuMF networks for each criterion, and (b) MONeuMF, which shares embeddings across criteria with multiple outputs.

can observe that individual NeuMF model was built based on each individual rating matrix associated with ratings in each criterion, in order to predict multi-criteria ratings, i.e., $\hat{R}_1, \hat{R}_2, ..., \hat{R}_K$.

While computationally straightforward, this independence assumption often oversimplifies real-world human cognition. However, these multi-criteria preferences are rarely isolated; for instance, a high rating for a hotel's location might influence its value rating. Or, a high rating on the ambience of the restaurant may also indicate a positive observation on dining services.

Dependent Predictions. To overcome the limitations of the independence assumption, recent advancements focus on dependent prediction models that explicitly capture inter-criteria correlations. A prominent example is the *Criteria Chains* model [81]. It organizes multiple criteria into a predefined sequence or "chain". Instead of predicting all criteria ratings simultaneously, it sequentially predicts a user's rating on each criterion one by one. During this process, the previously predicted criteria ratings are transformed and utilized as contextual information to help predict the user's preference for the subsequent criteria in the chain. It is classified as a dependent method because it explicitly captures and models these correlations. By feeding the predicted ratings of preceding criteria as direct inputs to estimate the subsequent ones, the model's predictions are mathematically dependent on prior outcomes in the chain.

Another example is multi-outputs NeuMF model which is a modern neural architecture [57], as shown by Figure 3b. Instead of building separate models for each criterion, MONeuMF utilizes a single, unified neural architecture where users and items share the same embedding layers. The network branches into multiple outputs, with each output neuron predicting the rating for a specific criterion. The entire model is optimized end-to-end using a joint loss function that aggregates the errors across all criteria. The joint loss function is actually a weighted sum of mean squared errors associated with predicted multi-criteria ratings, where the weights can be predefined as hyper-parameters. MONeuMF is classified as a dependent method because it implicitly captures these criteria dependencies. By forcing all criteria to share the same underlying user and item embeddings and optimizing them together via a joint loss function, the prediction error of one criterion directly influences and adjusts the shared representations. This joint learning mechanism ensures that the criteria are not treated in isolation, but rather as an interconnected system.

5.2. Stage 2: Preference Aggregation or Ranking

Following the prediction of multi-criteria ratings in Stage 1, the system proceeds to Stage 2, which aims to derive the final recommendation outcome. To provide a clear delineation of the underlying computational process, this stage encompasses two distinct procedural steps:

Step 1: Learning the Mapping Function (f). Before performing any aggregation on the test instances, the system must first learn a mapping function, denoted as f, using the historical training data. This function is designed to capture the relationship between a user's multi-criteria ratings and their final overall rating (or a derived ranking score). Depending on the architectural design, this mapping function can be either linear or non-linear. Traditional approaches frequently utilize linear aggregation (e.g., multiple linear regression [1]), assuming a straightforward weighted combination of criteria. In contrast, advanced models employ non-linear functions, such as support vector regression [82] or multi-layer perceptrons (used in both IndNeuMF [93] and MONeuMF [57] as shown in Figure 3), to capture more complex, high-dimensional interactions among different criteria.

The aggregation methods mentioned above, such as linear or non-linear regressions, usually were adopted to estimate the overall ratings. Alternatively, researchers may try to derive a ranking score to rank the candidate items to produce the top-N recommendations. This paradigm is referred to as *multi-criteria ranking*, which bypasses the explicit prediction of a single overall rating by directly evaluating the multidimensional preferences. For instance, recent studies have introduced ranking methods based on Pareto dominance, where a ranking score is derived directly from the dominance relationships among candidate items using their predicted multi-criteria ratings [93, 96, 97]. Drawing inspiration from multi-objective optimization techniques, such as Fonseca and Fleming's ranking, an item's final score is calculated based on the number of other candidate items it strictly dominates across all criteria. This approach ensures that items are ranked based on their absolute comparative advantage without relying on any aggregation loss function. Another example involves utility-based expectation models [85, 86]. Instead of a straightforward weighted combination, these methods first learn a user's specific expectations across multiple criteria. The system then estimates a utility or ranking score by calculating the similarity or distance between the user's learned expectations and the actual multi-criteria preferences for the candidate items. These ranking-oriented approaches offer a highly adaptable multi-criteria decision-making process, integrating with existing criteria rating prediction models to generate top-N recommendations.

Step 2: Preference Aggregation and Scoring. Once the mapping function f has been fully optimized and learned, it is applied to the unseen instances. The *predicted* multi-criteria ratings generated from Stage 1 are fed as inputs into this learned function f. The function then aggregates these predicted criteria values to calculate the estimated overall rating or a specific ranking score for the target user-item pair. By explicitly separating the learning of the aggregation function from the actual application of the predicted criteria ratings, the framework ensures a mathematically rigorous and highly adaptable multi-criteria decision-making process.

5.3. Multi-Stage MCRS: Connections to MCDM

Fundamentally, the multi-stage MCRS pipeline is closely linked to the classical theories of MCDM. In real-world scenarios, a user selecting an item acts as a decision-maker evaluating alternatives across various conflicting factors (e.g., location, cleanliness, and service in hotel bookings) [9, 67]. The aggregation and ranking processes in Stage 2 can be directly mapped to established MCDM paradigms, though their practical adoption in MCRS varies significantly due to computational and theoretical constraints.

5.3.1. Dominant Paradigms

The most universally adopted aggregation functions in MCRS correspond directly to the *Multi-Attribute Utility/Value Theory (MAUT)* in MCDM [41, 9]. Methods that utilize linear regressions, non-linear support vector regressions, or deep multi-layer perceptrons act as utility functions designed to aggregate multidimensional preferences into a single utility score (i.e., the overall rating). MAUT-based approaches are predominantly favored by MCRS researchers because they are straightforward to implement, easily optimized using historical overall ratings as ground truth, and integrated into end-to-end deep learning architectures.

Furthermore, the multi-criteria ranking approach based on Pareto dominance was inspired by MOO. It eliminates the need for mathematical aggregation functions, preserving the multidimensional integrity of user preferences until the final sorting step. However, it is necessary to clarify that this approach primarily leverages the theoretical concept of dominance to derive ranking scores, rather than translating the recommendation task into a strict MOO problem [93]. Formulating MCRS directly as a pure MOO task remains a profound challenge due to the inherently conflicting nature of user preferences. In classical MOO, the ultimate goal is to simultaneously optimize all given objectives or criteria. However, real-world users rarely maximize their satisfaction across all item aspects without making subjective compromises. For instance, a user seeking a hotel with exceptional ratings in room quality, service, and location must often compromise due to financial constraints, as such premium quality inevitably conflicts with a limited budget. Translating MCRS into a pure MOO task is difficult because existing recommendation datasets primarily capture the final evaluation ratings, lacking explicit records of the users' dynamic constraints (e.g., hard budget caps) and their underlying compromise strategies.

5.3.2. Underrepresented Paradigms

Despite a lot of paradigms in MCDM, several powerful techniques remain absent or rarely put into practice in multi-stage MCRS. Most notably, *Outranking models* (e.g., ELECTRE and PROMETHEE) [15] and advanced aggregation operators, such as TOPSIS [12] and Ordered Weighted Averaging (OWA) [24], are rarely utilized in modern multi-stage pipelines. While recent literature has begun exploring TOPSIS and OWA to handle the uncertainty and inconsistency of user preferences while preserving the relative importance of individual criteria, their application remains strictly confined to single-stage heuristic models. The reluctance to adopt these advanced MCDM methods in deep sequential pipelines stems from several critical structural and computational challenges.

One primary challenge is the requirement of explicit priors. Many advanced MCDM methods, such as TOPSIS and PROMETHEE, require explicit pre-defined weights or importance scores for each criterion from the decision-maker. In typical MCRS datasets, however, users only provide ratings rather than their explicit weight preferences.

This lack of explicit priors makes these methods difficult to apply without introducing additional heuristic weight-estimation steps or relying on objective entropy derivations.

Another significant hurdle lies in computational complexity and scalability. Outranking methods, for instance, rely heavily on extensive pairwise comparisons among all candidate alternatives across all criteria. In a modern recommender system with massive catalogs containing millions of items, performing such multidimensional pairwise comparisons for every user induces an immense computational overhead. This inherent complexity severely limits their applicability for generating real-time Top-N recommendations.

Beyond data and computational constraints, there exists an objective mismatch and architectural barrier. Traditional multi-stage models strictly optimize for regression metrics (e.g., RMSE or MAE) by predicting an absolute overall rating. In contrast, MCDM methods like TOPSIS inherently output a relative closeness coefficient designed exclusively for ranking rather than absolute value fitting. Furthermore, modern deep multi-stage architectures increasingly rely on end-to-end gradient descent and joint loss optimization. Classical MCDM operators involving sorting mechanisms (e.g., OWA) or ideal-solution geometric computations (e.g., TOPSIS) are typically non-differentiable, making their direct insertion into neural backpropagation pipelines mathematically non-trivial.

While theoretically sound, there remains a significant gap between classical MCDM theories and modern MCRS implementations. Many advanced MCDM models have simply not been adapted to handle the extreme data sparsity, high dimensionality, and dynamic nature typical of collaborative filtering environments. Bridging this gap by developing scalable, differentiable, and machine-learning-driven approximations of these advanced MCDM theories to enable true end-to-end multi-criteria ranking represents a highly promising, yet underexplored, frontier for future MCRS research.

6. Single-Stage MCRS: Holistic Modeling

By contrast, single-stage models adopt a holistic modeling paradigm. While multi-stage pipelines typically follow a fixed two-stage pattern of prediction and aggregation, single-stage MCRS does not adhere to a uniform architectural rule or fixed mechanism. Instead, its implementation heavily depends on how the underlying recommendation algorithms, such as latent factor models or probabilistic graphical models, are adapted to directly accommodate multi-dimensional criteria [94]. By treating the multiple criteria either as dimensions of similarity, latent probabilistic variables, or contextual features, single-stage models effectively mitigate the error propagation inherent in sequential pipelines. The following subsections detail the three predominant methodologies within this holistic framework.

6.1. Neighborhood-Based Approaches

Early single-stage models primarily extended traditional neighborhood-based collaborative filtering by redefining how user or item similarities are computed in a multi-dimensional space [1]. Multi-criteria ratings can be utilized to seek better neighbors to further be adopted in neighborhood-based collaborative filtering [44, 5]. More specifically, instead of relying solely on a single overall rating, these approaches compute similarities across each specific criterion independently. Subsequently, they combine these dimension-specific similarities into a unified global similarity metric using straightforward aggregation functions, such as calculating the arithmetic average or identifying the minimum similarity across all criteria. While these heuristic aggregations are computationally efficient, they frequently fail to capture the inherent uncertainty and conflicting nature of complex human preferences.

To overcome the limitations of simple linear aggregations, recent advancements have begun integrating sophisticated operations research operators directly into the memory-based collaborative filtering process. A prominent paradigm is the utilization of TOPSIS [12] and OWA [24] to aggregate multi-criteria ratings into a highly robust comprehensive score [8]. Rather than applying static weights to specific criteria, OWA sorts a user's multi-criteria ratings by magnitude and applies weights generated by fuzzy logic distribution functions. In classical operations research, these pre-defined mathematical functions are termed as "fuzzy linguistic quantifiers" representing concepts such as "at least half" or "as many as possible" purely through mathematical parameters rather than text mining. This systematically transforms the multi-dimensional ratings into a single aggregated score that intuitively reflects the user's general satisfaction level. Alternatively, TOPSIS utilizes data-driven objective weights, such as those derived from the CRITIC method, to evaluate how close a user's criteria ratings are to a hypothetical perfect evaluation and how far they are from the worst-case scenario. The resulting relative closeness coefficient serves as a robust aggregated rating. It is important to note that this aggregation occurs independently for each user-item pair. Thus, a user's entire profile of multi-criteria preferences is projected into a new, one-dimensional pseudo-rating vector (using either the ratings produced by OWA or the aggregated scores derived from TOPSIS) across all rated items. Using these newly constructed vectors, the system computes a multi-criteria similarity between users across their co-rated items via distance metrics. To establish the most accurate nearest neighbors, this implicitly derived multi-criteria similarity is then mathematically fused (e.g., via multiplication) with the traditional similarity calculated from the users' explicit overall ratings. By leveraging this similarity fusion, the system can identify a highly reliable neighborhood, enabling it to directly produce Top-N recommendations from highly sparse datasets without efforts to predict the unobserved criteria ratings.

In summary, the neighborhood-based methodologies discussed in this section encapsulate two distinct computational paradigms for leveraging multi-criteria preferences: computing similarities on individual criteria before aggregating them into a unified similarity metric, and alternatively, aggregating the multi-dimensional ratings into comprehensive pseudo-rating vectors prior to distance computation. Despite their procedural differences, both paradigms share the same objective: establishing a highly precise, multi-dimensional user-user similarity to identify better neighborhoods for collaborative filtering. Most importantly, by directly utilizing these refined neighborhoods to produce the final Top-N recommendations, both approaches inherently bypass the intermediate rating prediction of individual criteria for unobserved items. They exemplify a holistic, single-stage architectural paradigm by directly translating complex multi-criteria interactions into actionable recommendations.

6.2. Probabilistic Models

Beyond heuristic neighborhood calculations, probabilistic modeling offers a mathematically rigorous framework to capture the complex dependencies among multi-criteria ratings directly. An underlying challenge specific to multi-criteria rating data is the psychological phenomenon known as the halo effect. In real-world recommendation scenarios, a user's perception of one dominant attribute, or their pre-existing overall impression of an item, unconsciously biases their evaluations on other specific criteria [66]. Therefore, multi-criteria ratings are inherently highly correlated, meaning that assuming strict statistical independence among different criteria often leads to suboptimal modeling of user preferences.

To explicitly model these cognitive biases and underlying dependencies, researchers have successfully adapted probabilistic generative models, most notably the Flexible Mixture Model (FMM) [69]. The FMM operates by introducing latent class variables to cluster users and items, subsequently defining a joint probability distribution over users, items, and multi-dimensional ratings. By applying a structure learning algorithm, the model effectively incorporates inter-criteria dependencies into the probabilistic graphical network. Utilizing the Expectation-Maximization algorithm, the system infers the most likely overall preference score for unobserved user-item pairs purely through probabilistic reasoning over the joint distribution. This allows the system to effectively mitigate the halo effect and generate accurate recommendations without relying on rigid mathematical aggregation functions.

This FMM-based approach is classified as a single-stage architecture due to its capacity to completely bypass the intermediate rating prediction of individual criteria. Once the joint probability distribution and the network parameters are learned from the training data, the model can directly estimate the overall rating for any unseen user-item entry. During the inference phase, the system simply marginalizes out the unobserved multi-criteria variables, deriving the expected overall rating solely from the latent user and item class distributions [66]. The model does not require the multi-criteria ratings for a specific user-item pair to be known or explicitly predicted beforehand, embodying the holistic, one-step prediction paradigm of single-stage MCRS.

6.3. Latent Models

While probabilistic models successfully capture statistical dependencies, latent factor models are effective at uncovering hidden semantic features in highly sparse environments through advanced dimensionality reduction. In the single-stage multi-criteria context, the earliest breakthroughs extended traditional two-dimensional matrix factorization into Multilinear Singular Value Decomposition (MSVD) [48]. This approach constructs a three-dimensional tensor comprising users, items, and diverse rating criteria. By decomposing this multi-dimensional tensor into a shared low-dimensional latent space, MSVD uncovers both explicit and implicit relations among the three dimensions. Rather than calculating user similarities directly on the extremely sparse original multi-criteria rating matrix, the system identifies the active user's nearest neighbors by measuring distances within this dense, approximated latent tensor space. Once these highly accurate neighbors are identified, the algorithm reverts to a traditional memory-based collaborative filtering mechanism, utilizing the neighbors' historical records to generate the final Top-N recommendations. It is useful to note that, although the ultimate recommendation step employs a neighborhood-based collaborative filtering strategy, the principal mathematical contribution of this approach lies in the multilinear dimensionality reduction of the multi-criteria rating tensor. Thus, we still classify this methodology under latent models rather than heuristic neighborhood-based approaches.

Building upon this foundation, recent studies have further formalized the single-stage tensor paradigm by treating criteria as generalized contextual dimensions. For instance, Hong and Jung proposed a comprehensive multi-criteria tensor model that simultaneously incorporates users, items, criteria, and external context (e.g., cultural groups) into a single holistic tensor [35]. Instead of predicting multi-criteria ratings sequentially or independently, their approach preserves the inherent structural interrelations of the data. By applying higher-order SVD and stochastic gradient descent, the model directly factorizes the tensor to infer unobserved user preferences, exhibiting strong robustness to the severe data sparsity characteristic of multi-criteria environments.

Whether utilizing multilinear decomposition or contextual tensor factorization, these latent factor models embody the one-stage architectural philosophy. Rather than following a multi-stage pipeline that first predicts missing individual criteria ratings and subsequently aggregates them, these approaches map all users, items, and criteria into a unified, low-dimensional embedding space. By directly generating holistic user-item preference scores or similarities from this

Table 4
Recent Advances in Modern Single-Stage MCRS.

Models	Key Technology	Main Contributions
CPA-LGC [59]	Graph Learning	Criteria-aware graph representation learning
CA-GF [60]	Graph Filtering	Efficient graph-based recommendation
D-MGAC [26]	Graph + Attention + Contrastive Learning	Multi-view representation learning
TaTriGR [52]	MCDM-aware Graph Learning	Criterion weighting and ideal-point reasoning
HMCR [30]	Hyperbolic Learning	Hierarchical preference modeling
MCCR [31]	Causal Inference	Bias reduction and robust recommendation

shared latent space, they produce the final Top-N recommendations in a single end-to-end step, thereby avoiding the error propagation issues inherent in sequential prediction models.

6.4. Advances in Modern Single-Stage MCRS

MCRS has evolved rapidly in recent years driven by advances in graph representation learning, contrastive learning, causal inference, and more sophisticated preference modeling techniques, as described in Table 4. While early single-stage MCRS methods primarily relied on matrix factorization and tensor decomposition to learn latent user-item representations, recent studies increasingly leverage graph structures to capture complex interactions among users, items, and multiple criteria. Most recent innovations in MCRS have been concentrated within the Single-Stage paradigm, largely driven by advances in deep learning and representation learning.

6.4.1. MCRS Based on Graph Learning

One representative example is the development of graph-based MCRS. Recent studies have demonstrated that graph learning can effectively model higher-order relationships among users, items, and criteria while alleviating data sparsity issues. For example, CPA-LGC introduced a criteria preference-aware graph learning framework that explicitly models user-specific criterion preferences within a graph convolution architecture [59]. Building upon this line of research, CA-GF further explored graph filtering techniques and achieved competitive recommendation performance with substantially reduced computational complexity [60]. More recently, D-MGAC integrated multiview graph learning, dual-attention mechanisms, and contrastive learning strategies to jointly capture criterion-specific dependencies and improve representation quality [26]. Furthermore, TaTriGR proposed a targeting-aware tripartite graph learning framework that incorporates user-specific criterion weighting and ideal-point reasoning, representing one of the first attempts to more explicitly connect modern graph-based MCRS with principles from MCDM [52]. Although TaTriGR does not directly implement a classical MCDM method such as AHP or TOPSIS, it incorporates several MCDM principles, including user-specific criterion weighting, ideal-point reasoning, and lexicographic preference consistency. These mechanisms explicitly model criterion importance and trade-off relationships, thereby strengthening the connection between modern graph-based MCRS and decision-theoretic perspectives.

Despite the methodological differences among these graph-based approaches, they can all be categorized as Single-Stage MCRS because criteria information is integrated directly within a unified learning framework rather than through separate criterion prediction and subsequent aggregation processes. Unlike Multi-Stage MCRS, which explicitly predicts individual criterion ratings before deriving overall recommendations, these models learn latent representations and recommendation outcomes in an end-to-end manner. Accordingly, criterion interactions and dependencies can be captured implicitly during representation learning.

Compared with earlier Single-Stage approaches based on matrix factorization or tensor decomposition, graph-based MCRS offer several advantages. By explicitly modeling higher-order user–item–criterion relationships, graph structures can better alleviate data sparsity, capture complex dependencies among criteria, or exploit neighborhood information beyond pairwise interactions. However, these benefits come at the cost of reduced transparency and interpretability. Since criteria interactions are learned implicitly within deep graph representations, it is often more difficult to explain how individual criteria contribute to the final recommendation outcome. In addition, graph-based models generally require higher computational resources and more complex training procedures than traditional latent-factor approaches. Therefore, while modern graph-based MCRS may enhance predictive performance, they also introduce new challenges regarding explainability, scalability, and alignment with classical MCDM principles.

6.4.2. Hyperbolic Learning and Causal Inference

Beyond graph learning, several emerging directions have recently attracted attention. HMCR introduced hyperbolic representation learning into MCRS, enabling the modeling of hierarchical relationships among users, items, and criteria in non-Euclidean embedding spaces [30]. MCCR incorporated causal inference techniques into MCRS and demonstrated that explicitly modeling causal relationships can alleviate recommendation bias and improve recommendation robustness [31].

Although both HMCR and MCCR remain within the Single-Stage MCRS paradigm, they address limitations that are not adequately handled by traditional latent-factor or graph-based models. HMCR focuses on representation quality by exploiting the hierarchical structure of user preferences and criterion relationships, whereas MCCR focuses on improving recommendation reliability through causal reasoning and bias mitigation.

Compared with conventional graph-based MCRS, hyperbolic representation learning can more effectively capture hierarchical and multi-scale relationships in sparse recommendation environments. This may improve recommendation accuracy when user preferences exhibit complex criterion structures. However, hyperbolic models generally introduce additional mathematical complexity and may be more difficult to interpret and optimize than Euclidean embedding approaches. In contrast, MCCR does not primarily seek to improve representation learning but rather to address or alleviate recommendation bias. By explicitly modeling causal relationships, causal MCRS can potentially improve robustness, fairness, and generalization performance. Nevertheless, causal inference often requires stronger modeling assumptions and additional domain knowledge, and its effectiveness may depend on the availability and quality of causal signals.

Therefore, recent Single-Stage MCRS research is expanding beyond predictive accuracy toward broader objectives, including hierarchical preference modeling, robustness, bias reduction, and trustworthy recommendation. These developments illustrate how modern MCRS increasingly incorporates ideas from neighboring research areas while maintaining the same criterion integration mechanism.

These developments demonstrate that modern MCRS research has gradually expanded beyond traditional latent-factor approaches toward graph-based learning, causal reasoning, and hyperbolic representation learning. Nevertheless, despite the diversity of learning paradigms and modeling techniques, these advances primarily affect representation learning, preference modeling, and optimization strategies rather than altering the location at which criteria information is integrated. Therefore, they can still be naturally interpreted within the proposed taxonomy of Multi-Stage and Single-Stage MCRS architectures. In this sense, recent advances provide additional evidence supporting the explanatory power and continued relevance of the proposed taxonomy.

6.5. Single-Stage MCRS: Connections to MCDM

While single-stage architectures often originate from traditional collaborative filtering and representation learning, their mechanism of integrating multiple dimensions is aligned with classical MCDM paradigms. In recent literature, specific MCDM operators have been successfully integrated into single-stage pipelines to enhance holistic modeling. Most notably, aggregation-based MCDM techniques, such as TOPSIS and OWA, have been employed to aggregate multi-criteria ratings into a robust comprehensive score for computing global user similarities in memory-based models [8]. These applications illustrate that classical decision analysis can effectively address the ambiguity and conflict inherent in human judgments without relying on intermediate rating predictions, thereby integrating operations research principles into neighborhood-based formulations.

Despite these initial successes, a vast array of powerful MCDM methodologies remains conspicuously absent from single-stage MCRS literature. For instance, classical outranking models, such as ELECTRE and PROMETHEE, as well as preference disaggregation models, have rarely been adapted for holistic recommendation tasks [53]. However, these underutilized methods hold immense theoretical potential. Outranking models, which evaluate the absolute comparative advantages of items or users through strict dominance relationships, could revolutionize neighborhood-based approaches. Rather than relying on distance metrics or linear correlation measures to identify similar users, outranking methodologies define user neighborhoods through multidimensional preference dominance, providing a more flexible and robust collaborative filtering mechanism that naturally accommodates minor inconsistencies in user behavior.

The hesitation to fully integrate these advanced MCDM methods into single-stage architectures arises from a range of computational challenges. First, classical MCDM techniques traditionally require explicit pre-defined weights or importance scores for each criterion from the decision-maker. Because typical collaborative filtering datasets only contain sparse rating matrices without explicit weight preferences, integrating these methods necessitates additional complex, heuristic weight-estimation steps or objective entropy derivations. Second, methods like outranking rely heavily on extensive pairwise comparisons among all candidate alternatives across all criteria. In a modern recommender system with massive catalogs containing millions of items and users, performing such multidimensional pairwise comparisons induces an immense computational overhead, severely limiting their scalability for real-time Top-N recommendations [63]. Finally, traditional MCDM assumes a dense decision matrix, whereas MCRS environments are characterized by extreme data sparsity [53], making the direct application of classical decision analysis operators mathematically non-trivial and establishing a significant gap between MCDM theories and practical single-stage MCRS implementations.

7. Summary, Challenges and Futures

This section provides a reflective discussion on the major findings of this review. We first summarize the key characteristics, strengths, limitations, and applicable boundaries of different MCRS architectures from an MCDM perspective. We then discuss the remaining challenges that hinder the wider adoption of MCRS in real-world environments and outline several promising future research directions, including dynamic preference modeling, explainable recommendation, implicit criteria preference extraction, advanced

Table 5
Architectural Comparison between Multi-Stage and Single-Stage MCRS

Dimension	**Multi-Stage MCRS**	**Single-Stage MCRS**
Criteria Integration	Criteria ratings are explicitly predicted and subsequently aggregated.	Criteria information is directly incorporated into a unified recommendation framework.
Model Design	Modular architecture with separate prediction and aggregation components.	End-to-end architecture that jointly learns user, item, and criteria representations.
Method Reusability	Can directly reuse existing recommendation algorithms (e.g., CF, MF, NeuMF) and MCDM aggregation methods (e.g., MAUT, TOPSIS, OWA).	Often requires customized model architectures and joint optimization strategies.
Scalability	May suffer from increased complexity as the number of criteria grows due to multiple prediction tasks.	Generally more scalable from an architectural perspective because criteria are modeled jointly.
Computational Cost	Moderate training cost but potentially affected by maintaining multiple models and aggregation stages.	May incur substantial computational cost when using graph learning, contrastive learning, or deep neural architectures.
Error Propagation	Susceptible to error propagation because aggregation depends on predicted criteria ratings.	Avoids explicit error propagation by directly optimizing recommendation outcomes.
Criteria Dependency	Often treats criteria independently, although methods such as Criteria Chains attempt to model dependencies.	Naturally captures complex dependencies and interactions among criteria through joint learning.
Data Sparsity Robustness	Performance may deteriorate when individual criteria ratings are highly sparse.	Modern representation learning methods can better alleviate sparsity through shared embeddings and graph structures.
Compatibility with MCDM	Highly compatible with classical MCDM frameworks such as MAUT, OWA, TOPSIS, AHP, ELECTRE, and Pareto-based ranking.	MCDM concepts are usually learned implicitly within latent representations, making direct integration more challenging.
Decision Transparency	High. Intermediate criteria ratings and aggregation procedures are explicitly available for inspection.	Lower. Recommendation decisions are often generated through latent representations and complex optimization processes.
Explainability	Supports intrinsic model and decision transparency through explicit criteria evaluation and aggregation.	Typically relies on post-hoc explanation techniques such as feature attribution or attention analysis.
Typical Advantages	Flexible architecture, strong MCDM compatibility, transparent decision process, easy methodological reuse.	Strong predictive performance, superior dependency modeling, improved robustness to sparsity, end-to-end optimization.
Typical Limitations	Error propagation, weaker dependency modeling, scalability challenges with many criteria.	Higher development complexity, lower transparency, and potentially high computational requirements.
Best-Suited Scenarios	Decision-support applications where transparency, controllability, and MCDM integration are important.	Large-scale recommendation environments where predictive accuracy and representation learning are prioritized.

MCDM integration, and the incorporation of emerging AI technologies.

7.1. Architectures: Summary and Reflections

This review categorizes existing MCRS approaches into two architectural paradigms: Multi-Stage MCRS and Single-Stage MCRS. Unlike traditional taxonomies that primarily classify methods according to algorithmic families (e.g., collaborative filtering, matrix factorization, deep learning, or graph neural networks), the proposed taxonomy focuses on the role of criteria information within the recommendation process. The distinction is determined by whether criteria ratings are explicitly predicted and subsequently aggregated (i.e., Multi-Stage) or directly incorporated into a unified recommendation framework (i.e., Single-Stage). Table 5 summarizes the major differences between Multi-Stage and Single-Stage MCRS across multiple dimensions. Rather than viewing either architecture as universally superior, the comparison highlights their respective strengths, limitations, and applicable scenarios. More details are discussed in the following subsections.

7.1.1. Model Complexity and Scalability

Multi-Stage MCRS and Single-Stage MCRS exhibit distinct characteristics in model complexity. Multi-Stage architectures explicitly model individual criteria and subsequently aggregate criterion-level predictions into an overall recommendation. This modular design offers considerable flexibility, as different prediction models and aggregation mechanisms can be independently selected, optimized, and interpreted. For example, existing recommendation techniques such as collaborative filtering, matrix factorization, tensor factorization, and neural recommendation models can be directly employed to predict criterion ratings, while mature MCDM methods such as weighted utility aggregation, OWA operators, TOPSIS, AHP, and Pareto-based ranking can be adopted during the aggregation stage. As a result, Multi-Stage architectures often benefit from the reuse of well-established methodologies and do not necessarily require the development of highly specialized end-to-end recommendation models. However, one potential risk is error propagation, since the ultimate performance could be affected by multiple models and stages, such as the criteria rating prediction errors and overall preference aggregation errors.

However, such flexibility comes at the cost of increased architectural complexity. As the number of criteria grows, multiple prediction tasks must be maintained, trained, and updated, potentially increasing computational overhead and introducing error propagation across stages. Furthermore, independently optimizing criterion prediction and aggregation components may not always lead to globally optimal recommendation performance. The scalability of Multi-Stage architectures may become challenging in applications involving a large number of criteria or rapidly evolving user preferences.

In contrast, Single-Stage MCRS integrate criteria information directly into a unified learning framework. Early Single-Stage approaches relied primarily on probabilistic models, matrix factorization, and tensor decomposition, whereas recent advances increasingly leverage graph learning, contrastive learning, hyperbolic representation learning, and causal inference. By jointly modeling users, items, and criteria within a single optimization process, these approaches avoid explicit criterion prediction and aggregation stages. This end-to-end design often improves scalability and enables the modeling of complex interactions among criteria. Furthermore, recent representation learning techniques can effectively alleviate data sparsity and improve recommendation performance in large-scale environments.

Nevertheless, the scalability advantages of Single-Stage architectures do not necessarily imply lower computational cost. Modern graph-based and deep learning models may require substantial training resources, particularly when incorporating graph neural networks, contrastive learning, or causal reasoning modules. Moreover, designing unified architectures capable of simultaneously modeling users, items, criteria, and their interactions often requires more sophisticated model development and optimization strategies. Therefore, while Single-Stage MCRS generally offer stronger scalability from an architectural perspective, they may also introduce higher computational complexity and implementation challenges. This trade-off highlights the importance of considering both system requirements and application characteristics when selecting an appropriate MCRS architecture.

7.1.2. Criteria Dependency Modeling

Another important distinction between Multi-Stage and Single-Stage MCRS lies in their ability to model dependencies among criteria. In many real-world recommendation scenarios, decision criteria are not independent. For example, in restaurant recommendation, food quality and service quality may jointly influence user satisfaction, while in hotel recommendation, room cleanliness, location, and perceived value may exhibit complex interactions. Ignoring such relationships may limit the effectiveness of recommendation models.

Multi-Stage MCRS often treat individual criteria as separate prediction tasks and subsequently aggregate criterion-level predictions into an overall recommendation, though there have been attempts which try to capture these criteria dependencies, such as criteria chains. However, criterion dependencies are frequently modeled only implicitly or are entirely ignored. As a result, important interactions among criteria may not be fully captured, potentially limiting recommendation accuracy in domains where criteria exhibit strong correlations.

In contrast, Single-Stage MCRS are generally better suited for modeling criterion dependencies because criteria information is incorporated directly into a unified learning framework. Matrix factorization, tensor factorization, graph-based models, and modern representation learning approaches can jointly learn relationships among users, items, and multiple criteria. Recent graph-based MCRS further enhance this capability by explicitly modeling higher-order interactions and complex dependency structures. Thus, Single-Stage architectures often achieve superior performance in environments where recommendation outcomes are strongly influenced by interactions among criteria.

Nevertheless, stronger dependency modeling may also increase model complexity and reduce transparency. Furthermore, not all application domains require sophisticated dependency modeling. In domains where criteria are relatively independent or where interpretability and deployment simplicity are prioritized, Multi-Stage architectures may still provide a practical and effective solution. Therefore, the importance of criterion dependency modeling should be evaluated according to the characteristics of the target application rather than being viewed as a universal advantage.

7.1.3. Compatibility with MCDM

From an MCDM perspective, Multi-Stage MCRS exhibit a stronger conceptual alignment with classical decision-making processes. Traditional MCDM methods typically consist of several explicit steps, including criterion evaluation, criterion weighting, preference aggregation, and alternative ranking. Multi-Stage architectures naturally follow a similar workflow by first predicting criterion-level ratings and subsequently aggregating them into an overall recommendation. Many established MCDM techniques, such as weighted utility functions, OWA operators, TOPSIS, AHP, ELECTRE, and Pareto-based ranking, can be directly incorporated into the aggregation stage.

This compatibility provides several practical advantages. Existing MCDM methodologies can be reused without redesigning recommendation algorithms, enabling researchers and practitioners to leverage decades of research in decision analysis. Furthermore, the modular structure of Multi-Stage architectures allows different aggregation mechanisms to be easily compared, replaced, or customized according to application requirements and user preferences.

In contrast, Single-Stage MCRS generally place greater emphasis on representation learning and end-to-end optimization. Criterion weighting, preference aggregation, and ranking are often learned implicitly within a unified model rather than being explicitly defined. While this design can improve recommendation performance and capture complex interactions among criteria, it also makes the integration of traditional MCDM techniques more challenging. In many cases, classical MCDM concepts become embedded within latent representations rather than appearing as explicit decision-making components.

Recent developments nevertheless suggest increasing convergence between MCDM and modern MCRS. Emerging models such as TaTriGR incorporate concepts related to criterion weighting, ideal-point reasoning, and preference consistency into advanced graph-learning frameworks. These developments indicate that future MCRS research may move toward integrating MCRS models with MCDM architectures that combine the predictive power of modern representation learning with the decision-theoretic foundations of classical MCDM methods.

7.1.4. Explainability

From the perspective of model explainability [13, 16, 49] and decision explainability [7, 17, 68], Multi-Stage MCRS possess stronger intrinsic transparency because individual criteria ratings and aggregation mechanisms are explicitly modeled. The intermediate criterion-level predictions provide users and system designers with direct visibility into how recommendation outcomes are formed. Furthermore, many aggregation approaches adopted in Multi-Stage architectures, such as weighted utility functions, OWA operators, TOPSIS, AHP, and Pareto-based ranking, are grounded in well-established decision-theoretic principles. Therefore, the decision process itself is often easier to understand, analyze, and justify than that of black-box recommendation models.

For example, consider a restaurant recommendation scenario with three criteria: food quality, service quality, and price. A weighted aggregation model may predict criterion ratings of 4.8, 4.2, and 3.5, respectively, and compute the overall recommendation score using criterion weights of 0.5, 0.3, and 0.2. In this case, the model itself is transparent because both the criterion ratings and criterion importance can be directly examined, allowing users and analysts to understand how the final recommendation score is generated. Similarly, decision-oriented MCDM methods can provide intuitive explanations for recommendation outcomes. For instance, a TOPSIS-based recommender may explain that a restaurant is recommended because it is closest to the ideal alternative considering all criteria [46], whereas a Pareto-based approach may justify a recommendation by showing that the selected restaurant dominates competing alternatives in terms of food quality and service without being worse on other criteria. Such explanations are directly derived from the underlying decision process rather than from post-hoc interpretation techniques.

However, it is important to distinguish decision transparency from recommendation explainability [79, 76]. Although Multi-Stage architectures provide a more interpretable decision process, relatively few studies have systematically investigated whether such transparency translates into better user understanding, trust, satisfaction, or decision support effectiveness. One potential reason is that most publicly available MCRS datasets primarily contain ratings and criterion-level evaluations rather than user feedback regarding explanations. Consequently, researchers can readily evaluate prediction accuracy and ranking performance but often lack the infrastructure necessary to assess explanation quality through controlled user studies. In contrast to domains where researchers can easily deploy experimental systems and collect user interaction data, many MCRS applications rely on industry-generated datasets, making large-scale user-centered evaluation more difficult. As a result, the explainability of MCRS remains largely underexplored despite the inherent transparency provided by criterion-based recommendation processes. In practice, many MCRS studies implicitly assume that exposing criteria ratings and aggregation procedures is sufficient to explain recommendations, while rigorous user-centered evaluations of explanation quality remain limited. Recent work has begun to explore explainability in MCRS. For example, Rismala et al. [65] proposed a criteria-level explanation framework that generates natural-language explanations by matching user preferences and item aspects extracted from review texts. Rather than explaining the internal recommendation model, the approach focuses on explaining recommendation outcomes through criterion-specific aspect–opinion pairs. This line of research demonstrates the potential of leveraging multi-criteria information to generate more detailed recommendation explanations and improve transparency from the user's perspective. However, the authors also claimed that their study was limited since no user studies were performed.

Single-Stage MCRS typically learn criterion interactions implicitly through latent representations, graph structures, or deep neural architectures. Although such models often achieve superior predictive performance and stronger capability in modeling complex criterion dependencies, their decision processes are generally less transparent. As a result, explanations frequently rely on post-hoc interpretation techniques, such as feature attribution, attention analysis, or model-agnostic explanation methods. This distinction highlights an important trade-off between decision transparency and predictive capability when selecting appropriate MCRS architectures. It also suggests that explainability remains an underexplored research direction in MCRS, particularly from the perspective of user-centered evaluation and decision support.

As a summary, neither architectural paradigm is universally superior. Multi-Stage MCRS are particularly suitable for applications that require explainability, explicit decision support, and direct integration with MCDM methodologies. In contrast, Single-Stage MCRS are often preferable in large-scale recommendation environments where predictive performance, scalability, and the modeling of complex criterion interactions are prioritized. Therefore, the choice between Multi-Stage and Single-Stage architectures should ultimately depend on the characteristics of the application domain, the availability of criteria information, and the desired balance between interpretability and predictive accuracy.

Recent advances further demonstrate the continued relevance of the proposed taxonomy. Although modern MCRS increasingly incorporate graph neural networks, attention mechanisms, contrastive learning, causal reasoning, and other advanced AI techniques, these developments primarily modify representation learning and optimization strategies rather than changing how criteria information is integrated into the recommendation pipeline. Recent state-of-the-art models can still be naturally interpreted within the Multi-Stage versus Single-Stage framework, providing additional evidence supporting the explanatory power and longevity of the proposed taxonomy.

7.2. Challenges and Future Directions

Although the integration of MCDM theories and advanced machine learning models has accelerated the evolution of recommender systems, a substantial gap persists between theoretical MCDM frameworks and practical MCRS implementations. Bridging this gap requires the research community to make a transition from purely algorithmic optimizations toward more holistic, dynamic, and human-centric paradigms.

7.2.1. Human-Centric Evaluations for MCRS

A primary challenge lies in the prevailing evaluation methodology of MCRS, which heavily relies on offline predictive metrics such as RMSE or MAE in rating predictions or precision and recall in top-N recommendations. Classical MCDM emphasizes the cognitive processes and interactions of human decision-makers; however, contemporary MCRS research rarely incorporates comprehensive user-centric studies. Because multi-criteria preferences are inherently subjective, users often adopt cognitive compromises when evaluating conflicting criteria, a phenomenon well-documented in operations research but largely ignored in recommendation algorithms. Future research should prioritize a shift toward human-centric MCRS, investigating how factors such as trust, emotional states, and personality traits influence multi-dimensional decision-making [94].

7.2.2. Modeling Dynamic Preferences in MCRS

Beyond human-centric evaluations, the dynamic nature of user preferences presents a significant challenge. Current MCRS architectures predominantly assume that a user's multi-criteria ratings reflect an absolute, static goodness. In reality, a user's perception of specific criteria is highly dynamic upon dynamic contexts. For instance, a traveler's evaluation of a hotel's "room size" or "location" may vary drastically depending on whether they are on a business trip or a family vacation. Integrating context-awareness with MCRS constitutes a highly promising research trajectory [91, 23, 4]. Developing advanced models that treat criteria preferences as dynamic variables conditioned by external contextual factors (e.g., time, companion, and occasion), rather than static side information, may substantially improve the precision of personalized recommendations.

7.2.3. Integration with MCDM

On the algorithmic front, several classical MCDM methodologies have already been successfully adapted to enhance multi-criteria recommender systems. For instance, multi-attribute utility functions [94], aggregation operators like OWA and TOPSIS [8], and Pareto dominance [93] derived from MOO have demonstrated remarkable effectiveness in modeling user preferences and generating recommendations in MCRS. Despite these successes, a vast array of other powerful MCDM frameworks, such as outranking models (e.g., ELECTRE and PROMETHEE) and preference disaggregation models, remain conspicuously underutilized in the MCRS literature [94]. The primary challenge to integrate these advanced methods stems from their inherent structural and computational limitations. Traditional MCDM methods were originally designed to evaluate a limited number of alternatives and criteria within small-scale, dense decision matrices. When redeployed into modern MCRS environments involving extreme data sparsity and millions of users and items, they face severe scalability bottlenecks and prohibitive computation costs. Methodologies like outranking rely heavily on establishing strict preference relationships, which necessitates exhaustive multidimensional pairwise comparisons among all candidate alternatives. Nevertheless, these formidable computational challenges may be progressively resolved alongside the continuous advancement of hardware and software technologies. A prime example of this technological mitigation can be observed in the recent application of Pareto dominance for multi-criteria ranking. While constructing Pareto-based rankings traditionally suffers from the immense computational overhead of exhaustive pairwise comparisons, researchers have successfully circumvented this bottleneck by deploying advanced parallel computing techniques, specifically utilizing NVIDIA GPU-based CUDA programming to drastically accelerate the calculations and make the process highly efficient [93, 98]. This successful acceleration paradigm suggests that the continuous evolution of parallel computing could serve as the key to eventually unlocking and reusing other computationally expensive MCDM theories for large-scale MCRS in the future.

7.2.4. Explainable MCRS

Explainability represents another promising research direction for future MCRS development. Although the explicit modeling of multiple criteria naturally provides richer information for explanation generation than traditional single-rating recommender systems, explainability remains relatively underexplored in the MCRS literature. Relatively little attention has been devoted to understanding how recommendation models generate criterion-level predictions and how individual criteria contribute to final recommendation decisions. From an MCDM perspective, MCRS offers unique opportunities for explainability because recommendation decisions can potentially be decomposed into criterion evaluation, criterion weighting, and preference aggregation processes. In particular, Multi-Stage MCRS naturally support decision transparency by exposing intermediate criterion ratings and aggregation mechanisms. Nevertheless, it remains unclear whether such transparency translates into improved user understanding, trust, satisfaction, or decision quality. One reason is that explainability evaluation in MCRS has largely focused on explanation coverage and relevance, while user-centered evaluations remain limited. Future research should therefore investigate explainability from both algorithmic and human-centered perspectives. Potential directions include developing criteria-aware explanation frameworks, integrating explainable AI techniques into modern graph-based and neural MCRS, generating natural-language explanations for criterion interactions and trade-offs, and conducting user studies to evaluate the effectiveness of explanations in supporting decision making. The integration of explainable recommendation techniques with established MCDM concepts may further strengthen the transparency, trustworthiness, and practical applicability of MCRS in decision-support environments.

7.2.5. Handling Missing or Partial Preferences

A pervasive, yet insufficiently addressed, issue in current MCRS literature is the missing criteria rating problem. Most existing algorithms operate under the unrealistic assumption that the training set contains complete evaluations across all criteria for every observed user-item pair. In reality, explicitly collecting ratings across every dimension is time-consuming and frequently faces user reluctance. To mitigate user fatigue, real-world platforms like TripAdvisor now often randomly sample only a subset of criteria for a user to evaluate, which results in generating partial data [99]. Moreover, while recent approaches attempt to extract implicit multi-criteria preferences directly from textual reviews, users' natural language feedback rarely covers every predefined criterion [99]. Consequently, the extracted criteria matrices inevitably contain massive missing values. Developing more robust algorithms and imputation mechanisms capable of handling partially observed criteria ratings without degrading the recommendation accuracy remains a direction for future development.

To further address extreme data sparsity and missing values, the field must embrace more innovative representation learning paradigms. Currently, while there is a massive surge of research integrating Large Language Models (LLMs) with general recommender systems, their specific adaptation and service for MCRS remain conspicuously scarce. The future points toward customizing LLMs to intelligently extract implicit multi-criteria preferences from unstructured textual reviews, which can significantly alleviate the user fatigue associated with explicit numerical ratings while handling the complex semantics of multi-dimensional opinions [99, 63]. Leveraging these emerging technologies will not only mitigate existing computational bottlenecks but also broaden the applicability of MCRS to more general environments.

7.2.6. Fusing MCRS with Other RSs

Another promising research direction is the fusion of MCRS with other recommendation paradigms. Although MCRS has traditionally focused on modeling users' preferences across multiple criteria, many real-world recommendation environments involve additional dimensions such as contextual situations, group decision making, conversational interactions, and multiple stakeholders. The integration of these paradigms may further improve the applicability and effectiveness of MCRS in complex decision-making scenarios.

Existing studies have demonstrated the benefits of combining multi-criteria preferences with contextual information, where users may assign different importance to criteria under different contextual situations [91, 23, 4, 47]. Similarly, criterion-level information has been incorporated into group recommender systems to better capture heterogeneous preferences among group members and improve group satisfaction estimation [20, 84, 47]. Beyond context and groups, multi-stakeholder recommendation environments introduce additional challenges because recommendation decisions must balance the utilities of multiple stakeholders whose objectives may be partially aligned or conflicting [89, 92, 85]. These studies suggest that multi-criteria preferences can provide a useful representation for modeling richer decision-making processes beyond traditional single-user recommendation scenarios.

Recent advances in conversational recommendation and large language models further create opportunities for dynamically eliciting criterion preferences, explaining recommendation trade-offs, and supporting interactive decision making. Rather than viewing MCRS as an isolated recommendation paradigm, future research may investigate unified frameworks that jointly model criteria, contexts, groups, stakeholders, and conversational interactions. Such integrations may substantially broaden the applicability of MCRS while strengthening its connection to real-world decision-support environments.

8. Conclusion

In this narrative review, we have systematically bridged the theoretical foundations of classical MCDM with the modern algorithmic implementations of MCRS. This paper provides a conceptually-driven taxonomy that categorizes the expansive MCRS literature into two architectures: multi-stage and single-stage models. Multi-stage architectures explicitly decompose the recommendation pipeline into criteria rating prediction and preference aggregation, heavily relying on operations research principles such as multi-attribute utility functions and Pareto dominance. Conversely, single-stage architectures embody a holistic, end-to-end modeling philosophy, leveraging advanced representation learning techniques, ranging from tensor factorization to multiview graph neural networks, to directly translate complex multi-criteria interactions into actionable recommendations without intermediate predictions.

Furthermore, this review highlights a critical paradigm shift in the current MCRS landscape: the transition from relying on explicitly elicited numerical ratings to mining implicit multi-dimensional preferences directly from user-generated textual reviews. This pivotal shift not only mitigates the pervasive issues of data sparsity and user fatigue but also significantly broadens the real-world applicability of multi-criteria recommendation algorithms across diverse domains. Despite these profound technical advancements, our comprehensive analysis reveals that a substantial gap persists between theoretical decision science and large-scale recommendation engineering. Key challenges for the community include scaling computationally intensive MCDM methods to massive sparse datasets, integrating context-aware dynamics, and advancing human-centric evaluations that address fairness, accountability, and transparency. Ultimately, by integrating classical operations research with advanced machine learning, this review aims to offer a solid theoretical foundation and a clear roadmap to guide the development of accurate, explainable, and human-centric multi-criteria recommender systems.

CRediT authorship contribution statement

Yong Zheng: Conceptualization of this study, Manuscript writing, reviewing and editing. **David Xuejun Wang:** Conceptualization of this study, Manuscript writing, reviewing and editing.

Declaration of competing interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of usage of GenAI. During the preparation of this work the authors used ChatGPT in order to improve language and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding acknowledgement. Not available.

Data availability. Data availability has been declared in Section 2.3.

References

[1] Adomavicius, G., Kwon, Y., 2007. New recommendation techniques for multicriteria rating systems. IEEE Intelligent Systems 22, 48–55.

[2] Adomavicius, G., Manouselis, N., Kwon, Y., 2010. Multi-criteria recommender systems, in: Recommender systems handbook. Springer, pp. 769–803.

[3] Adomavicius, G., Tuzhilin, A., 2011. Context-aware recommender systems, in: Recommender systems handbook. Springer, pp. 217–253.

[4] Afzal, I., Yilmazel, B., Kaleli, C., 2024. An approach for multi-context-aware multi-criteria recommender systems based on deep learning. IEEE Access 12, 99936–99948.

[5] Aggarwal, C.C., 2016. Neighborhood-based collaborative filtering, in: Recommender Systems: The Textbook. Springer, pp. 29–70.

[6] Al-Ghuribi, S.M., Noah, S.A.M., 2019. Multi-criteria review-based recommender system–the state of the art. IEEE Access 7, 169446–169468.

[7] Alufaisan, Y., Marusich, L.R., Bakdash, J.Z., Zhou, Y., Kantarcioglu, M., 2021. Does explainable artificial intelligence improve human decision-making?, in: Proceedings of the AAAI conference on artificial intelligence, pp. 6618–6626.

[8] Anwar, K., Wasid, M., Zafar, A., Iqbal, A., 2025. An efficient framework to combat multidimensionality in multi-criteria recommender systems. Cluster Computing 28, 245.

[9] Aruldoss, M., Lakshmi, T.M., Venkatesan, V.P., 2013. A survey on multi criteria decision making methods and its applications. American Journal of Information Systems 1, 31–43.

[10] Bączkiewicz, A., Kizielewicz, B., Shekhovtsov, A., Wątróbski, J., Więckowski, J., Salabun, W., 2021. Towards an e-commerce recommendation system based on mcdm methods, in: 2021 International Conference on Decision Aid Sciences and Application (DASA), IEEE. pp. 991–996.

[11] Batmaz, Z., Kaleli, C., 2019. Ae-mccf: an autoencoder-based multi-criteria recommendation algorithm. Arabian Journal for Science and Engineering 44, 9235–9247.

[12] Behzadian, M., Otaghsara, S.K., Yazdani, M., Ignatius, J., 2012. A state-of the-art survey of topsis applications. Expert Systems with applications 39, 13051–13069.

[13] Belle, V., Papantonis, I., 2021. Principles and practice of explainable machine learning. Frontiers in big Data 4, 688969.

[14] Bouyssou, D., 2000. Evaluation and decision models: a critical perspective. volume 32. Springer Science & Business Media.

[15] Bouyssou, D., 2009. Outranking methods. Encyclopedia of optimization 4, 249–255.

[16] Burkart, N., Huber, M.F., 2021. A survey on the explainability of supervised machine learning. Journal of Artificial Intelligence Research 70, 245–317.

[17] Coussement, K., Abedin, M.Z., Kraus, M., Maldonado, S., Topuz, K., 2024. Explainable ai for enhanced decision-making. Decision Support Systems 184, 114276.

[18] Dara, S., Chowdary, C.R., Kumar, C., 2020. A survey on group recommender systems. Journal of Intelligent Information Systems 54, 271–295.

[19] Deb, K., Sindhya, K., Hakanen, J., 2016. Multi-objective optimization, in: Decision Sciences. CRC Press, pp. 161–200.

[20] Dixit, V.S., Mehta, H., Bedi, P., 2014. A proposed framework for group-based multi-criteria recommendations. Applied Artificial Intelligence 28, 917–956.

[21] Doumpos, M., Zopounidis, C., 2011. Preference disaggregation and statistical learning for multicriteria decision support: A review. European Journal of Operational Research 209, 203–214.

[22] Doumpos, M., Zopounidis, C., 2019. Preference disaggregation for multicriteria decision aiding: An overview and perspectives. New Perspectives in Multiple Criteria Decision Making: Innovative Applications and Case Studies , 115–130.

[23] Dridi, R., Tamine, L., Slimani, Y., 2022. Exploiting context-awareness and multi-criteria decision making to improve items recommendation using a tripartite graph-based model. Information Processing & Management 59, 102861.

[24] Emrouznejad, A., Marra, M., 2014. Ordered weighted averaging operators 1988–2014: A citation-based literature survey. International Journal of Intelligent Systems 29, 994–1014.

[25] Figueira, J., Greco, S., Ehrgott, M., 2005. Multiple criteria decision analysis: state of the art surveys .

[26] Forouzandeh, S., Krivitsky, P.N., Chandra, R., 2025. Multiview graph dual-attention deep learning and contrastive learning for multi-criteria recommender systems. Expert Systems with Applications 291, 128378.

[27] Forouzandeh, S., Rostami, M., Berahmand, K., 2022. A hybrid method for recommendation systems based on tourism with an evolutionary algorithm and topsis model. Fuzzy Information and Engineering 14, 26–50.

[28] Fuchs, M., Zanker, M., 2012. Multi-criteria ratings for recommender systems: an empirical analysis in the tourism domain, in: International conference on electronic commerce and web technologies, Springer. pp. 100–111.

[29] Greco, S., Figueira, J., Ehrgott, M., 2016. Multiple criteria decision analysis. volume 37. Springer.

[30] Guo, Z., Han, T., Song, P., Feng, C., Yao, K., Liang, J., 2025a. Hyperbolic multi-criteria rating recommendation, in: Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2162–2171.

[31] Guo, Z., Song, P., Feng, C., Yao, K., Liang, J., 2025b. Causal inference for multi-criteria rating recommender systems. ACM Transactions on Information Systems 43, 1–30.

[32] Hariri, N., Mobasher, B., Burke, R., Zheng, Y., 2011. Context-aware recommendation based on review mining., in: ITWP Workshop at IJCAI.

[33] Hassan, M., Hamada, M., 2017. A neural networks approach for improving the accuracy of multi-criteria recommender systems. Applied Sciences 7, 868.

[34] Hassan, M., Hamada, M., 2019. Evaluating the performance of a neural network-based multi-criteria recommender system. International Journal of Spatio-Temporal Data Science 1, 54–69.

[35] Hong, M., Jung, J.J., 2021. Multi-criteria tensor model for tourism recommender systems. Expert Systems with Applications 170, 114537.

[36] Huang, J., Meng, Y., Guo, F., Ji, H., Han, J., 2020. Weakly-supervised aspect-based sentiment analysis via joint aspect-sentiment topic embedding , 6989–6999.

[37] Ishizaka, A., Nemery, P., 2013. Multi-criteria decision analysis: methods and software. John Wiley & Sons.

[38] Jacquet-Lagreze, E., Siskos, J., 1982. Assessing a set of additive utility functions for multicriteria decision-making, the uta method. European journal of operational research 10, 151–164.

[39] Jacquet-Lagreze, E., Siskos, Y., 2001. Preference disaggregation: 20 years of mcda experience. European Journal of Operational Research 130, 233–245.

[40] Jannach, D., Zanker, M., Fuchs, M., 2014. Leveraging multi-criteria customer feedback for satisfaction analysis and improved recommendations. Information Technology & Tourism 14, 119–149.

[41] Jansen, S.J., 2011. The multi-attribute utility method, in: The measurement and analysis of housing preference and choice. Springer, pp. 101–125.

[42] Jia, W., Shan, J., Wang, J., Mengyang, Z., 2026. Implicit aspect-based sentiment analysis: A systematic review. ACM Computing Surveys 58, 1–37.

[43] Khan, M.M., Ibrahim, R., Ghani, I., 2017. Cross domain recommender systems: A systematic literature review. ACM Computing Surveys (CSUR) 50, 1–34.

[44] Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R., Riedl, J., 1997. Grouplens: Applying collaborative filtering to usenet news. Communications of the ACM 40, 77–87.

[45] Kumar, R., Pamucar, D., 2025. A comprehensive and systematic review of multi-criteria decision-making (mcdm) methods to solve decision-making problems: two decades from 2004 to 2024. Spectrum of Decision Making and Applications 2, 177–196.

[46] Kumbum, P.K., Adari, V.K., Chunduru, V.K., Gonepally, S., Amuda, K.K., 2020. Artificial intelligence using topsis method. International Journal of Research Publications in Engineering, Technology and Management 3, 4305–4311.

[47] Le, N.L., Abel, M.H., 2025. Integrating context and criteria: a multi-head attention-based approach for multi-criteria group recommender systems. World Wide Web 28, 50.

[48] Li, Q., Wang, C., Geng, G., 2008. Improving personalized services in mobile commerce by a novel multicriteria rating approach, in: Proceedings of the 17th international conference on World Wide Web, pp. 1235–1236.

[49] Lisboa, P.J., Saralajew, S., Vellido, A., Fernández-Domenech, R., Villmann, T., 2023. The coming of age of interpretable and explainable machine learning models. Neurocomputing 535, 25–39.

[50] Masthoff, J., 2015. Group recommender systems: aggregation, satisfaction and group attributes, in: recommender systems handbook. Springer, pp. 743–776.

[51] Mateos, P., Bellogín, A., 2024. A systematic literature review of recent advances on context-aware recommender systems. Artificial Intelligence Review 58, 20.

[52] Meng, Z., Meng, F., 2026. From criteria to ranking: Targeting-aware tripartite graph learning for multi-criteria recommendation, in: Proceedings of the ACM Web Conference 2026, pp. 6421–6432.

[53] Monti, D., Rizzo, G., Morisio, M., 2021. A systematic literature review of multicriteria recommender systems. Artificial intelligence review 54, 427–468.

[54] Moshkovich, H., Mechitov, A., Olson, D., 2016. Verbal decision analysis, in: Multiple Criteria Decision Analysis: State of the Art Surveys. Springer, pp. 605–636.

[55] Murtadha, A., Wen, B., Pan, S., Su, J., Ao, L., Liu, Y., 2024. Bert-asc: Auxiliary-sentence construction for implicit aspect learning in sentiment analysis. Expert Systems with Applications 258, 125195.

[56] Musto, C., De Gemmis, M., Semeraro, G., Lops, P., 2017. A multi-criteria recommender system exploiting aspect-based sentiment analysis of users' reviews, in: Proceedings of the eleventh ACM conference on recommender systems, pp. 321–325.

[57] Nassar, N., Jafar, A., Rahhal, Y., 2020. Multi-criteria collaborative filtering recommender by fusing deep neural network and matrix factorization. Journal of Big Data 7, 1–12.

[58] Ouatik, F., Erritali, M., Ouatik, F., Jourhmane, M., 2022. Student orientation recommender system using topsis and ahp. Journal of Information and Organizational Sciences 46, 473–489.

[59] Park, J.D., Li, S., Cao, X., Shin, W.Y., 2023. Criteria tell you more than ratings: Criteria preference-aware light graph convolution for effective multi-criteria recommendation, in: Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining, pp. 1808–1819.

[60] Park, J.D., Yoo, J., Shin, W.Y., 2025. Criteria-aware graph filtering: Extremely fast yet accurate multi-criteria recommendation, in: Proceedings of the ACM on Web Conference 2025, pp. 4482–4493.

[61] Pelissari, R., Alencar, P.S., Amor, S.B., Duarte, L.T., 2022. The use of multiple criteria decision aiding methods in recommender systems: A literature review, in: Brazilian Conference on Intelligent Systems, Springer. pp. 535–549.

[62] Peters, M.D., Godfrey, C.M., Khalil, H., McInerney, P., Parker, D., Soares, C.B., 2015. Guidance for conducting systematic scoping reviews. JBI Evidence Implementation 13, 141–146.

[63] Rajput, I.S., Tewari, A.S., Tiwari, A.K., 2025. A prisma-based systematic review on deep learning in multi-criteria recommender systems. Procedia Computer Science 259, 1533–1542.

[64] Rismala, R., Maulidevi, N.U., Surendro, K., 2023. Multi-attribute bert for preferences completion in multi-criteria recommender system, in: Proceedings of the 2023 12th International Conference on Software and Computer Applications, pp. 315–320.

[65] Rismala, R., Maulidevi, N.U., Surendro, K., 2025. Enhancing the comprehensiveness of criteria-level explanation in multi-criteria recommender system. Journal of Information Systems Engineering & Business Intelligence 11.

[66] Sahoo, N., Krishnan, R., Duncan, G., Callan, J., 2012. Research note—the halo effect in multicomponent ratings and its implications for recommender systems: The case of yahoo! movies. Information Systems Research 23, 231–246.

[67] Sahoo, S.K., Goswami, S.S., 2023. A comprehensive review of multiple criteria decision-making (mcdm) methods: advancements, applications, and future directions. Decision making advances 1, 25–48.

[68] Shi, C., Yao, Y., 2025. Explainable multi-criteria decision-making: a three-way decision perspective. International Journal of Approximate Reasoning , 109528.

[69] Si, L., Jin, R., 2003. Flexible mixture model for collaborative filtering, in: Proceedings of the 20th International Conference on Machine Learning (ICML-03), pp. 704–711.

[70] Snyder, H., 2019. Literature review as a research methodology: An overview and guidelines. Journal of business research 104, 333–339.

[71] Stefanidi, E., Bentvelzen, M., Woźniak, P.W., Kosch, T., Woźniak, M.P., Mildner, T., Schneegass, S., Müller, H., Niess, J., 2023. Literature reviews in hci: A review of reviews, in: Proceedings of the 2023 CHI conference on human factors in computing systems, pp. 1–24.

[72] Torra, V., 2011. Learning Aggregation Operators for Preference Modeling. Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 317–333.

[73] Wang, H., Lu, Y., Zhai, C., 2010. Latent aspect rating analysis on review text data: a rating regression approach, in: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 783–792.

[74] Wang, H., Lu, Y., Zhai, C., 2011. Latent aspect rating analysis without aspect keyword supervision, in: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 618–626.

[75] Wang, H., Olhofer, M., Jin, Y., 2017. A mini-review on preference modeling and articulation in multi-objective optimization: current status and challenges. Complex & Intelligent Systems 3, 233–245.

[76] Wardatzky, K., Inel, O., Rossetto, L., Bernstein, A., 2025. Whom do explanations serve? a systematic literature survey of user characteristics in explainable recommender systems evaluation. ACM Transactions on Recommender Systems 3, 1–35.

[77] Wongvilaisakul, W., Netinant, P., Rukhiran, M., 2023. Dynamic multi-criteria decision making of graduate admission recommender system: Ahp and fuzzy ahp approaches. Sustainability 15, 9758.

[78] Zhang, K., Zhang, K., Zhang, M., Zhao, H., Liu, Q., Wu, W., Chen, E., 2022. Incorporating dynamic semantics into pre-trained language model for aspect-based sentiment analysis , 3599–3610.

[79] Zhang, Y., Chen, X., 2020. Explainable recommendation: A survey and new perspectives. Foundations and Trends® in Information Retrieval 14, 1–101.

[80] Zhang, Y., Lai, G., Zhang, M., Zhang, Y., Liu, Y., Ma, S., 2014. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis, in: Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, pp. 83–92.

[81] Zheng, Y., 2017a. Criteria chains: a novel multi-criteria recommendation approach, in: Proceedings of the 22nd International Conference on Intelligent User Interfaces, pp. 29–33.

[82] Zheng, Y., 2017b. Situation-aware multi-criteria recommender system: using criteria preferences as contexts, in: Proceedings of the symposium on applied computing, pp. 689–692.

[83] Zheng, Y., 2018. Personality-aware decision making in educational learning, in: Companion Proceedings of the 23rd International Conference on Intelligent User Interfaces, pp. 1–2.

[84] Zheng, Y., 2019a. Educational group recommendations by learning group expectations, in: 2019 IEEE International Conference on Engineering, Technology and Education (TALE), IEEE. pp. 1–7.

[85] Zheng, Y., 2019b. Utility-based multi-criteria recommender systems, in: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, pp. 2529–2531.

[86] Zheng, Y., 2020. Penalty-enhanced utility-based multi-criteria recommendations. Information 11, 551.

[87] Zheng, Y., 2023. ITM-Rec: An open data set for educational recommender systems, in: Companion Proceedings of the 13th International Conference on Learning Analytics & Knowledge (LAK).

[88] Zheng, Y., 2025. Opentable data with multi-criteria ratings. Published on IEEE DataPort: `https://doi.org/10.21227/8avw-yk62`.

[89] Zheng, Y., Ghane, N., Sabouri, M., 2019a. Personalized educational learning with multi-stakeholder optimizations, in: Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization, pp. 283–289.

[90] Zheng, Y., Mobasher, B., 2018. Context-Aware Recommendations. World Scientific Publishing. pp. 173–202.

[91] Zheng, Y., Shekhar, S., Jose, A.A., Rai, S.K., 2019b. Integrating context-awareness and multi-criteria decision making in educational learning, in: Proceedings of the 34th ACM/SIGAPP symposium on applied computing, pp. 2453–2460.

[92] Zheng, Y., Toribio, J.R., 2021. The role of transparency in multi-stakeholder educational recommendations: Y. zheng, jr toribio. User modeling and user-adapted interaction 31, 513–540.

[93] Zheng, Y., Wang, D., 2022a. Multi-criteria ranking: Next generation of multi-criteria recommendation framework. IEEE Access 10, 90715–90725.

[94] Zheng, Y., Wang, D., 2023a. Multi-criteria decision making and recommender systems, in: Companion Proceedings of the 28th International Conference on Intelligent User Interfaces, pp. 181–184.

[95] Zheng, Y., Wang, D.X., 2022b. A survey of recommender systems with multi-objective optimization. Neurocomputing 474, 141–153.

[96] Zheng, Y., Wang, D.X., 2023b. A comparative study of preference ordering methods for multi-criteria ranking, in: 2023 10th IEEE Swiss Conference on Data Science (SDS), IEEE. pp. 108–111.

[97] Zheng, Y., Wang, D.X., 2023c. Multi-criteria ranking by using relaxed pareto ranking methods, in: Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization, pp. 81–85.

[98] Zheng, Y., Wang, D.X., Ruan, Q., 2024. MCRecKit: An open-source library for multi-criteria recommendations, in: 2024 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), IEEE. pp. 892–897.

[99] Zhuang, Y., Kim, J., 2021. A bert-based multi-criteria recommender system for hotel promotion management. Sustainability 13, 8039.