Empirical and Experimental Perspectives on Big Data in Recommendation Systems: A Comprehensive Survey

Kamal Taha*, Paul D. Yoo, Chan Yeun, and Aya Taha

Abstract: This survey paper provides a comprehensive analysis of big data algorithms in recommendation systems, addressing the lack of depth and precision in existing literature. It proposes a two-pronged approach: a thorough analysis of current algorithms and a novel, hierarchical taxonomy for precise categorization. The taxonomy is based on a tri-level hierarchy, starting with the methodology category and narrowing down to specific techniques. Such a framework allows for a structured and comprehensive classification of algorithms, assisting researchers in understanding the interrelationships among diverse algorithms and techniques. Covering a wide range of algorithms, this taxonomy first categorizes algorithms into four main analysis types: user and item similarity based methods, hybrid and combined approaches, deep learning and algorithmic methods, and mathematical modeling methods, with further subdivisions into sub-categories and techniques. The paper incorporates both empirical and experimental evaluations to differentiate between the techniques. The empirical evaluation ranks the techniques based on four criteria. The experimental assessments rank the algorithms that belong to the same category, sub-category, technique, and sub-technique. Also, the paper illuminates the future prospects of big data techniques in recommendation systems, underscoring potential advancements and opportunities for further research in this fields.

Key words: big data algorithms; recommendation systems; recommendation algorithms; deep learning in recommendations

1 Introduction

The rapid advancement of big data has fundamentally transformed the analytical landscape, reshaping interactions and decision-making processes for both corporations and consumers[1, 2]. This transformation is fueled by the immense potential of big data, which enables users to explore a broad spectrum of questions, including unforeseen ones[3]. In today's digital age, personal mobile devices transcend their traditional role as communication tools; they are now potent data collectors, capturing every facet of an individual's digital footprint, from communications to online transactions. Corporations benefit from extensive data collection, gaining insights into individual and group preferences[4].

This torrent of data brings forth a significant challenge: the increasing complexity for users to make decisions that align with their personal needs[5]. Similarly, corporations grapple with the monumental

• Kamal Taha and Chan Yeun are with Department of Computer Science, Khalifa University, Abu Dhabi 127788, United Arab Emirates (UAE). E-mail: kamal.taha@ku.ac.ae; chan.yeun@ku.ac.ae.

• Paul D. Yoo is with School of Computing and Mathematical Sciences, Birkbeck College, University of London, London WC1E 7HU, UK. E-mail: p.yoo@bbk.ac.uk.

• Aya Taha is with Department of Science, Brighton College, Dubai 122002, United Arab Emirates. E-mail: bc004814@brightoncollegedubai.ae.

* To whom correspondence should be addressed.

task of processing and interpreting this data to anticipate user behaviors[6]. In response, recommendation systems have emerged as a pivotal solution within the realm of machine learning[7]. These systems use algorithms to analyze past user activities and their similarities with others, predicting future interests to provide personalized recommendations.

Recommendation systems have proliferated across various sectors, including Internet of Things (IoTs) services, entertainment, e-learning, web search, bioinformatics, and engineering[8–11]. Their operational mechanisms range from content-based filtering, which builds user profiles based on item characteristics to recommend similar items, to collaborative filtering, relying on a user's and others' past behaviors, and to advanced deep learning based systems that utilize AI algorithms to parse extensive datasets across platforms, adeptly personalizing user experiences[12]. Hybrid systems often yield more precise recommendations.

The role of data in shaping smart cities is also paramount[13–15]. This information is vital for smart city development, government planning, and enhancing individual comfort. Major companies like Amazon, Google, and Microsoft have effectively employed hybrid recommendation technologies to deliver personalized products and news.

Research in the field of recommendation systems can be traced back to the 1990s[16]. During this era, numerous heuristics were developed for content-based and Collaborative Filtering (CF) methodologies[17]. Gaining significant attention with the Netflix challenge, Matrix Factorization (MF) emerged as the predominant model in recommender systems from 2008 to 2016[18, 19]. However, the linear characteristics of factorization models limited their effectiveness in handling large datasets, such as intricate user-item interactions and items with complex semantics[20, 21].

Coinciding with these developments, the mid-2010s witnessed a surge in the application of deep neural networks within machine learning, significantly impacting fields such as speech recognition, computer vision, and natural language processing[22]. This trend has also extended to graph models, where innovative multi-task prompting methods, inspired by prompt learning in Natural Language Processing (NLP), have shown potential in enhancing the adaptability and effectiveness of graph-based machine learning applications[23].

The success of deep learning is attributed to the extensive expressiveness of neural networks, which are especially beneficial for learning from large datasets with complex patterns[24]. This advancement offers new prospects for enhancing recommendation technologies. Consequently, in recent years, there has been a notable increase in research focusing on neural network approaches to recommender systems[25]. Moreover, the integration of sociological behavioral criteria with data mining techniques, as noted in Ref. [26], highlights the complexity of user interactions in digital environments. Utilizing advanced models such as hypergraphs and neural networks, this approach is key to creating sophisticated and precise recommendation systems.

While several surveys have examined various facets of recommendation systems, this comprehensive survey aims to delve deeper into the current state of big data in recommendation systems, incorporating both experimental and empirical evaluations. Our goal is to provide an in-depth and nuanced understanding of this field, highlighting its achievements and challenges, thereby paving the way for future advancements.

1.1 Motivation and key contributions

(1) Main challenge and proposed solution

(a) **Current issue:** The current surveys examining algorithms in recommendation systems utilizing big data are deficient in two primary respects: first, there is an absence of a comprehensive and current overview; second, the classification of algorithms is excessively general and lacks precision[17, 27–30]. This shortfall leads to ambiguity in categorizing diverse algorithms and results in the application of uniform metrics, which yields inaccurate evaluations.

(b) **Proposed solution:** We propose a twofold approach: first, we conduct an in-depth analysis of current algorithms for big data in recommendation systems; second, we introduce a methodologically sound taxonomy. This taxonomy offers a hierarchical and detailed classification of these algorithms, fostering a more accurate and systematic categorization.

(2) Comprehensive survey and enhanced assessment

(a) **Survey goals:** We provide a survey of algorithms, focusing on those that use the same categories, sub-categories, techniques, and sub-categories.

(b) Benefits of the taxonomy: Employing our taxonomy enhances the precision of assessments and comparisons of algorithms. It offers insights into their strengths and weaknesses, aiding future research.

(3) Empirical and experimental evaluations

(a) Empirical evaluation: The study includes an empirical evaluation of diverse techniques used in big data for recommendation systems.

(b) Experimental evaluation: This study ranks the algorithms that utilize same categories, sub-categories, techniques, and sub-categories experimentally.

We classify the algorithms for recommendation systems with big data into four broad categories based on their " analysis methods". These categories are: (1) user and item similarity based methods; (2) hybrid and combined approaches; (3) deep learning and algorithmic, and (4) mathematical modeling methods. Each of these primary categories is divided into more

specific sub-categories. These sub-categories are then broken down into various techniques that utilize the principles of their respective sub-category. Lastly, these techniques are further categorized into more detailed sub-techniques. Figure 1 shows our methodology-based taxonomy. Our taxonomy offers the following benefits:

• Structured presentation: This approach offers a systematic framework for displaying survey findings. The arrangement of related methodologies within a hierarchical layout aids in understanding the paper's logical progression.

• Extensive inclusivity: This classification system encompasses a broad range of relevant techniques. Its tiered structure assists in pinpointing unexplored areas and research opportunities.

• Technique analysis: By categorizing comparable methods together, this taxonomy facilitates the
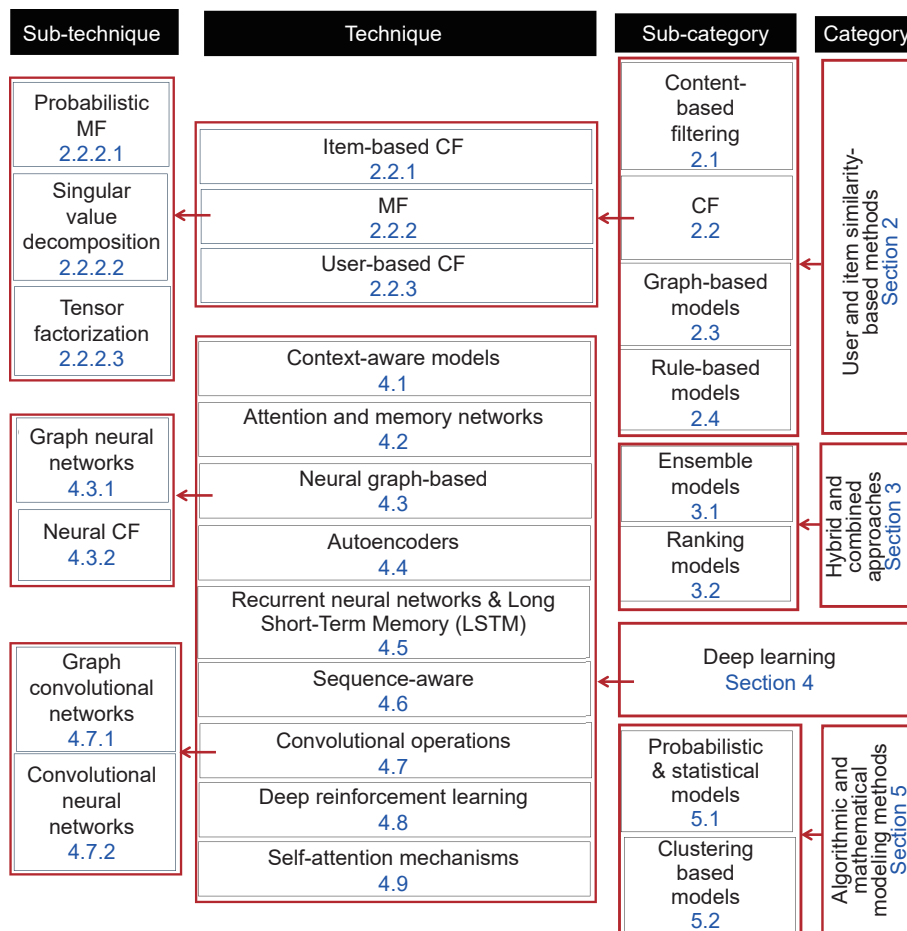

Fig. 1 Structured and hierarchical approach to categorizing algorithms used in big data for recommendation systems. It methodically breaks down the algorithms into detailed categories, moving from broader methodology categories to more specific sub-techniques. The categorization in the top layer is based on analysis methods. The figure provides the corresponding section numbers in the manuscript.

comparison of various research techniques. It highlights both their commonalities and distinctions, enabling an evaluation of their respective merits and limitations.

• Enhanced replicability: This categorization improves the ability to replicate research by providing clear descriptions of the methodologies.

2 User and Item Similarity Based Method

2.1 Content-based filtering

This approach to recommendations combines detailed analysis of item features with user preferences to provide personalized suggestions. Initially, it builds a user profile based on their interactions and preferences. This profile is then used to compare against the attributes of various items, which can be diverse and abundant in big data contexts, ranging from genre, keywords, and specific features relevant to the items, to more complex data like text descriptions, categories, images, or audio features.

Items are recommended by scoring each one based on its similarity to the user's profile. For example, if a user shows a preference for movies by a certain director, the system will recommend other movies by that same director. The effectiveness of this approach is heightened when there is detailed, descriptive data about each item, allowing for a more accurate match to the user's preferences. This technique utilizes advanced data processing and machine learning methods, such as natural language processing and computer vision, to handle big data's complexity. This ensures recommendations are personalized, relevant, and accurate, thus enhancing user experience by aligning suggestions closely with individual interests.

The rationale behind the usage of this technique: (a) Personalization: It focuses on the preferences of individual users, which leads to a high level of personalization; (b) Item independence: It does not rely on other users' behavior, which is particularly beneficial in situations where there is sparse data on user interactions; (c) Transparency: Recommendations can be easily explained by item features; (d) New item inclusion: It can recommend new items that have no previous user interactions; (e) Handling big data: it can efficiently manage vast and diverse item features, making it a scalable solution; and (f) Stability: The recommendations are stable over time as they are based

on the user's consistent preferences. This is useful for users with niche or specific tastes.

The conditions for the optimal performance of this technique: (a) Effective processing of textual data through advanced NLP techniques is crucial. This involves understanding context, semantics, and nuances in language; (b) Employing robust machine learning algorithms that can learn from data, adapt to new information, and make accurate predictions is essential; (c) Tailoring recommendations to users, rather than making broad, generic suggestions. This involves understanding unique user preferences; and (d) Implementing a feedback loop where the system learns from the user's reactions to its recommendations. This helps in refining accuracy.

The limitations of this technique: (a) The approach leads to a lack of diversity in recommendations. It tends to recommend items similar to those a user has liked in the past, leading to a filter bubble where the user is not exposed to a broader range of options; (b) For new users, there's a lack of historical data to base recommendations on. This makes it difficult to predict preferences; (c) New items without any user interactions or ratings can be challenging to recommend, as the system has no data to assess their relevance to users; (d) Its effectiveness is dependent on the quality of the metadata associated with the items; and (e) It may struggle with subjective user preferences due to its difficulty in capturing subtleties.

Table 1 evaluates and discusses research papers that have utilized content-based filtering for big data in recommendation systems.

2.2 Collaborative filtering

2.2.1 Item-based collaborative filtering

The technique of item-based recommendation focuses on recommending items to users by analyzing the similarity between items, utilizing metrics such as cosine similarity, Pearson correlation, or adjusted cosine similarity. This approach contrasts with user-similarity methods, as it compares items based on user-item interaction data like ratings or viewing history. This process helps in understanding how different items relate in terms of user preferences. For instance, if a user highly rates a movie, the system recommends other movies like it based on these similarity metrics.

Item-based recommendation excels in large environments due to its scalability and efficiency, particularly in handling sparse data, resulting in more

Table 1   Research papers that have employed content-based filtering for big data in recommendation systems.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Hu et al.[27] 2020	Amazon Camera review	Fair	Acceptable	Acceptable	The authors created CISER, a system merging user profiling, reviewer credibility, and sentiment analysis. It has five parts: (1) Extracting key features using context and sentiment; (2) Scoring reviewers based on expertise and trust; (3) Mining user interests from review styles; (4) Assigning ratings to features based on sentiment; and (5) Recommending products using credibility-weighted sentiment.
Wang et al.[28] 2022	DBLP and APS	Acceptable	Unsatisfactory	Good	The authors introduced CNCRec, a citation recommendation framework that combines collaborative filtering and network representation learning. It is tailored for diverse academic networks, creating a paper rating matrix from an attributed citation network, incorporating paper attributes, such as topics from titles and abstracts. CNCRec optimizes citation representation by uniting the citation network and paper attributes.
Jiang et al.[29] 2016	Travelogues and photos	Good	Acceptable	Unsatisfactory	The authors developed a system for personalized travel sequence recommendations, leveraging travelogues and user-uploaded photos with tags and location information. It prioritizes entire travel sequences over single sites, using a "topical package space" with details like tags, cost, time, and season. The system matches and suggests travel routes using user similar users' travel histories.
Lian et al.[30] 2018	LBSN dataset	Fair	Fair	Good	The authors introduced ICCF, a framework for implicit feedback collaborative filtering, which improves parameter learning using coordinate descent. They combined ICCF with graph Laplacian matrix factorization to demonstrate the effectiveness of user features in assessing mobility similarity for location recommendations in a large location-based social network dataset.
Liu et al.[31] 2019	Author collected	Acceptable	Fair	Good	The authors developed a system to help undergraduates locate and work on research projects. It analyzes students' data, activities, and networks for relevance, connectivity, and quality. Utilizing data mining, social network analysis, and bibliometric analysis, the system recommends projects that align with students' qualifications.

consistent and accurate suggestions. With fewer items than users and more stable item relationships, it is easier to manage and effective in big data contexts.

The rationale behind the usage of this technique: (a) Item-based collaborative filtering scales better in big data settings, as item-item matrices are smaller than user-based ones, aiding in managing resources; (b) It typically offers more precise recommendations by analyzing detailed item relationships; (c) This method efficiently addresses the sparsity in user-item interactions common in big data; (d) Its recommendations are more transparent, being easier to understand due to item similarities; and (e) It effectively accommodates new users by suggesting popular items, independent of their past interactions.

The conditions for the optimal performance of this technique: Efficient item similarity calculation using methods like cosine similarity, Pearson correlation, or Jaccard index is key, chosen based on data type. In big data, enhancing sparse item-user matrices with dimensionality reduction or data imputation improves recommendation quality. Scalability, critical for handling growing data, is often managed with distributed computing frameworks, such as Apache Hadoop or Spark. Balancing personalization and diversity in recommendations ensures both similarity and variety. For new items/users with minimal interaction history, content-based filtering or hybrid models are effective.

The limitations of this technique: Scalability

challenges in large datasets require more processing power and memory for similarity calculations, risking performance issues. Sparse user-item interactions in big data often lead to inaccurate recommendations. New items or users lack historical data, hindering accurate recommendation. Frequent recommendation of popular items overshadows niche ones, reducing diversity. Changes in user preferences/item features make recommendations from item CF outdated.

Table 2 evaluates and discusses research papers that have utilized Item-Based Collaborative Filtering for Big Data in recommendation systems.

2.2.2 Matrix factorization

2.2.2.1 Probabilistic Matrix Factorization (PMF)

The technique employs a probabilistic approach to decompose large user-item interaction matrices, which typically represent user-item interactions such as user ratings for movies on a streaming platform. Each entry in these matrices corresponds to a user's rating or implicit feedback for a specific item. This probabilistic nature is crucial for handling the uncertainty and noise inherent in real-world data, as it assumes that these interactions are influenced by latent (hidden) factors related to both users and items, like genres, movie length, and user preferences. By learning latent features for both users and items and assigning probability distributions to these features instead of fixed values, the technique is able to make personalized and accurate recommendations even with sparse and noisy data. It tackles sparse data by predicting missing matrix entries, allowing for user rating predictions on non-interacted items. Its flexibility and scalability make it effective for handling vast and intricate datasets, ensuring efficient, personalized recommendations.

The rationale behind the usage of this technique: PMF excels in handling large, sparse user-item matrices common in big data, efficiently processing millions of users and items. It extracts meaningful latent features, scaling well for big data applications. PMF's probabilistic approach models uncertainty in user preferences and item attributes, useful in noisy, incomplete data scenarios. By reducing data dimensionality and capturing latent user-item relationships, PMF often outperforms simpler methods.

The conditions for the optimal performance of this technique: Optimizing PMF in recommendation

Table 2   Research papers that have employed item-based collaborative filtering for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Yan et al.[32] 2018	WS-DREAM	Unsatisfactory	Good	Good	The authors improved the Item-based Collaborative Filtering (ICF) method by adding the LSH technique, aiming for secure data publication and efficient integration in distributed service recommendations. Their proposed ICFLSH method assists recommender systems in offering services while maintaining user privacy.
Dai et al.[10] 2020	MovieLens and Netflix	Unsatisfactory	Unsatisfactory	Good	The authors developed CoFiToR, a three-stage transfer learning-based collaborative filtering framework that refines user preferences from general to specific, enhancing top-N recommendations. They also introduced a logistics recommendation method encompassing user tracking, choice evaluation, matrix-based rating analysis, habitual behavior reflection, and user information analysis via a model and decoder.
Sun and Zhang[33] 2022	GroupLens MovieLens	Unsatisfactory	Fair	Acceptable	The authors introduced a modified fuzzy adaptive resonance theory-based biclustering method for user similarity, focusing on shared items. The approach combines biclustering for local similarity and prediction with item-based collaborative filtering for global predictions, culminating in a merged, comprehensive final prediction.
Zan[34] 2023	MovieLens	Unsatisfactory	Fair	Fair	The authors introduced a logistics service recommendation technique involving user interaction tracking, influenced choice evaluation, matrix-based rating analysis, habitual behavior reflection with a time interval matrix, and user information analysis through a model and decoder.

systems is anchored in large, high-quality datasets for accurate user-item interaction learning. Selecting an appropriate number of latent features is essential to avoid underfitting or overfitting and manage computational complexity. Critical to this process is the tuning of hyperparameters such as learning rate and regularization. Also, enhancing PMF with other methods can significantly boost its effectiveness.

The limitations of this technique: PMF in large recommendation systems struggles with sparse data and faces scalability issues as datasets grow, demanding more computational resources. Its linear approach may not capture complex user-item relationships as effectively as non-linear models. PMF is challenged by new users or items without sufficient historical data. Its performance depends heavily on precise hyperparameter tuning, with a risk of overfitting if feature dimensions are too high or regularization is not well-adjusted. The interpretability of PMF's latent factors is limited.

Table 3 evaluates and discusses research papers that have utilized PMF for big data in recommendation systems.

2.2.2.2 Singular Value Decomposition (SVD)

SVD starts by representing the data as a matrix, where rows could be users and columns could be items. This matrix is often sparse as not all users interact with all items. SVD helps to fill in these missing values by approximating the original matrix. The decomposition results in three matrices: a diagonal matrix of singular values and two orthogonal matrices. These singular values are key to understanding the strength of the latent features in the data. By truncating these values, SVD can reduce the dimensionality of the data, which helps in handling big datasets more efficiently. This reduced representation is then used to predict missing entries in the original matrix,

The rationale behind the usage of this techniques: (a) SVD identifies latent features in data, like genres in movies, improving accuracy in predicting user preferences; (b) SVD effectively handles sparse data in recommendation systems by estimating missing values, aiding in accurate predictions even with limited data; (c) Advances in algorithms and computing allow SVD to efficiently process large datasets, making it practical for big data applications; (d) SVD improves

Table 3   Research papers that employed PMF for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Yi et al.[35] 2019	MovieLens, Douban-Book	Fair	Unsatisfactory	Good	The authors introduced a deep learning collaborative filtering approach, deep matrix factorization, incorporating auxiliary information. It applies two transformation functions to create latent factors for users and items from different inputs. They created an embedding for recommendation systems that predicts positive feedback and condenses sparse, high-dimensional data into compact, low-dimensional vectors.
Hu et al.[36] 2022	Movielens, Jester	Unsatisfactory	Good	Good	The authors created RAP, a privacy-focused recommender system framework. It mixes private user ratings with public ones using local disturbance, allowing algorithms to use modified ratings without accessing private ones. Users employ de-perturbation for recommendations.
Yao et al.[37] 2021	Programmable Web	Unsatisfactory	Acceptable	Good	The authors created a mashup service recommendation system by adding implicit API correlation regularization to matrix factorization. They emphasize the importance of API characteristics and past interactions with mashups for future API predictions. They detailed the model's components and proposed methods to understand their approach's functionality.
Shen et al.[38] 2021	MovieLens, GroupLens	Unsatisfactory	Fair	Good	The authors developed DVMF, a Bayesian recommendation framework using variational inference for enhanced optimization. The Parametric Inference Model, with dual neural networks, generates hyperparameters for latent factors, leading to the variational MF model. The method merges implicit feedback and user/item data into one-hot format using implicit feedback.

recommendation quality by revealing subtle user preferences not obvious in basic analyses; (e) Integral to collaborative filtering, SVD predicts user preferences for non-interacted items based on user-item interactions; and (f) Despite its complexity, SVD's output is interpretable, offering transparency and aiding in system refinement.

The conditions for the optimal performance of this technique: (a) Regularization is essential for controlling overfitting in sparse datasets, it helps SVD models generalize better; (b) Implementations must support parallel processing and distributed computing for large datasets; (c) Tuning parameters, like the number of latent factors, using methods like cross-validation, is crucial for SVD's effectiveness; (d) Developing strategies for new users or items with limited data is vital for system consistency; (e) SVD models should accommodate changing user preferences and item popularity over time; and (f) Ensuring the model's robustness to missing data through effective imputation techniques is important for accuracy.

The limitations of this technique: (a) SVD is ineffective with new users or items lacking historical data; (b) SVD may fail to capture complex, non-linear data relationships; (c) Performance heavily relies on correctly setting hyperparameters like latent factors; (d) SVD is prone to overfitting with datasets having numerous features; (e) Extracted latent features by SVD are often difficult to interpret; (f) The model's performance can degrade with noisy data; and (g) SVD does not naturally incorporate user-specific biases or preferences.

Table 4 evaluates and discusses research papers that have utilized SVD for big data in recommendation systems.

2.2.2.3 Tensor factorization

The technique employs tensor factorization, a process that involves decomposing a high-dimensional tensor into multiple, lower-dimensional matrices or tensors. A tensor is essentially a multi-dimensional array of data. For example, in the context of a movie recommendation system, a 3D tensor might encompass dimensions for users, movies, and time. This setup represents how users' preferences for movies evolve over time.

Tensor factorization simplifies large, sparse tensors into smaller, dense ones, similar to dividing a complex puzzle into easier pieces. Techniques like higher-order singular value decomposition or CANDECOMP/PARAFAC decomposition are used for this purpose. By reducing data dimensions, it makes data more manageable, focusing on retaining the most crucial features.

This approach is particularly crucial in big data contexts, where the volume of data can be overwhelming. Factorization helps to uncover latent (hidden) factors that influence user preferences and behaviors. For instance, in a movie recommendation system, these factors might represent underlying genres or themes that explain why certain users prefer certain

Table 4   Research papers that have employed SVD for big data in recommendations.

Pape/year	Dataset	Interpretability	Scalability	Efficiency	Description
Guan et al.[39] 2023	Ciao, Epinions, Flixster	Fair	Acceptable	Acceptable	The authors introduced the Simultaneous Community detection and Singular Value Decomposition (SCSVD) framework, enhancing recommender systems via community detection. SCSVD combines community detection and recommendation model creation, linking ratings and social networks. It iteratively optimizes user preferences using social network communities, improving preference.
Lian et al.[40] 2021	Yelp and Amazon	Unsatisfactory	Fair	Good	The authors created a recommender system framework using discrete matrix factorization, accommodating diverse datasets and loss functions with explicit and implicit feedback, and auxiliary data. It uses block coordinate descent and semidefinite relaxation, functioning without hyperparameters in a two-phase item recall and ranking process.
Jiao et al.[41] 2020	Movielens	Unsatisfactory	Acceptable	Acceptable	The authors proposed an Adaptive Learning Rate (ALR) function combining exponential and linear elements, integrated into the SVD++ recommendation algorithm. Featuring a high initial value, the learning rate quickly decreases in the middle phase and then slowly reduces to a smaller value at the end.

movies. The goal is to discover these latent factors that explain observed data, thereby simplifying and enhancing the understanding of complex data.

The rationale behind the technique: Tensor factorization enhances recommendation systems by handling multi-dimensional real-world data (like user, item, and time), capturing complex patterns beyond traditional two-dimensional matrix factorization. This approach is especially beneficial in big data scenarios, where it can identify intricate relationships and offer more accurate predictions.

The conditions for optimal performance: Big data's high dimensionality and sparsity are addressed by Principal Components Analysis (PCA) for dimension reduction and matrix completion for sparsity. Choosing correct hyperparameters like factors, regularization, and learning rates is essential for model success. Model complexity must be balanced to capture data patterns without overfitting, using cross-validation and regularization. Including contextual and temporal elements enhances performance. Testing various tensor factorization algorithms, such as CP and Tucker decomposition, finds the best fit for specific cases.

The limitations of this technique: Tensor factorization struggles with scalability in large datasets due to increased computational complexity and memory needs. Sparsity in data, characterized by many missing tensor entries, hampers learning effective representations and predictions. It also faces the cold start problem, making it difficult to accurately recommend for new users or items with sparse data.

Table 5 evaluates and discusses research papers that have utilized tensor factorization for big data in recommendation systems.

2.2.3 User-based collaborative filtering

The technique in question involves creating detailed user profiles based on past behaviors and preferences, a key component of big data environments. This method operates on the assumption that users with similar tastes or behaviors in the past will likely continue to exhibit similar preferences in the future. To facilitate this, large datasets of user interactions and preferences, including ratings, views, purchases, or other forms of engagement with products, movies, or articles, are collected and analyzed. The technique uses metrics like cosine similarity to match user profiles with similar preferences, predicting a user's future likes based on others' tastes, enhancing recommendation accuracy.

The rationale behind the usage of this technique: It can effectively personalize recommendations by analyzing the behavior and preferences of similar users. It effectively uses implicit data, such as browsing history, to infer preferences, particularly when explicit feedback like ratings is limited. It diversifies recommendations by leveraging the varied tastes of a broad user base. It incorporates social relationships and community influences to tailor suggestions. It is effective in sparse big data environments.

The conditions for the optimal performance of this technique: (a) In big data, sparse user-item matrices are common. Using MF, dimensionality reduction, or adding more data sources, like content-based filtering, helps address this; (b) Recommendation systems need to balance personalization with diversity to prevent filter bubbles, thus broadening user exposure and fostering new interests; and (c) Incorporating both explicit (ratings) and implicit (click-through rates) user feedback is key for continually improving recommendation algorithms.

The limitations are: (a) It struggles with scalability in large systems, causing performance issues; (b) Sparse interactions in big datasets hinder finding similar users, reducing recommendation accuracy; (c) New users or items with limited history pose data insufficiency problems; (d) A bias towards recommending popular items often overlooks niche content; and (e) It tends to reinforce content homogenization, limiting user experience diversity.

Table 6 evaluates and discusses research papers that have utilized user-based collaborative filtering for big data in recommendation systems.

2.3 Graph-based models

The approach involves leveraging graph structures to model and analyze complex relationships in large datasets, where data is represented as a graph consisting of nodes and edges. Nodes typically represent items, such as products, movies, or articles, and users, while edges signify relationships or interactions like purchases, ratings, or views. The primary concept here is the utilization of graph theory to capture and analyze these relationships, making it easier to explore connections, detect patterns, and make predictions based on the nature and strength of these connections.

Enhancing this approach, the use of graph algorithms

Table 5　**Featuring research papers that have employed tensor factorization for big data in recommendations.**

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Ioannidis et al.[42] 2021	Digg dataset	Acceptable	Acceptable	Good	The authors proposed Coupled Graph-Tensor Factorization (CGTF), integrating graph-related auxiliary data into recommender systems. They designed an algorithm for factor matrix determination and missing entry completion. CGTF employs an ADMM solver for closed-form updates, supporting parallel and accelerated execution. This model also tackles the cold-start problem, occurring when tensors have incomplete slabs.
Li et al.[43] 2018	QoS dataset	Unsatisfactory	Unsatisfactory	Good	The authors introduced a time-aware matrix factorization model for cloud service QoS prediction, combining adaptive matrix factorization with temporal smoothing to provide accurate, time-sensitive forecasts for service recommendations. Utilizes temporal smoothing for more precise, timely QoS predictions in recommendations.
Meng et al.[44] 2018	QoS dataset	Acceptable	Good	Acceptable	The authors presented a service recommendation method combining time and location using tensor factorization. It uses CP decomposition for multi-dimensional analysis. It distinguishes between varying and stable QoS metrics, addresses data scarcity, and uses CP decomposition for user-service-location relationships.
Yang et al.[45] 2018	Last.fm, Bibsonomy	Fair	Fair	Acceptable	The authors created Tagrec-CMTF, a tag recommendation system using CMTF, which leverages auxiliary matrices and tensor CP factorization to uncover latent features and optimize learning in tags, items, and users.

Table 6　**Research papers that have employed user-based collaborative filtering for big data in recommendations.**

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Qi et al.[46] 2018	MovieLens	Unsatisfactory	Unsatisfactory	Good	SBT-Rec is a method developed by the authors, utilizing structural balance theory. It identifies a target user's "enemies" (users with contrasting preferences) in E-commerce, then finds "possible friends" (considering "enemy's enemy as a friend"). It recommends products liked by these "possible friends," which also match the user's preferences.
X. Wu and Z. Wu[47] 2023	Author collected	Fair	Fair	Acceptable	The authors introduced a system for recommending products, utilizing a CF algorithm and extensive data search capabilities. This system employs its filtering algorithm to sift through vast datasets during data processing, enhancing efficiency and optimizing the performance of the recommendation system.
Hong and Jung[48] 2018	Author collected	Unsatisfactory	Unsatisfactory	Fair	The authors proposed a method for recommending content to groups using social media data, considering social influence, emotional contagion, and conformity. They use an emotion-based model to identify emotional connections among users, which informs their group recommendations. Then, they are incorporated into group recommendations.

such as shortest path, clustering, or graph neural networks enables the analysis of graphs to uncover patterns in interconnected data, crucial for accurate recommendations. These algorithms analyze paths and connections, allowing the system to provide personalized recommendations based on user preferences and item connections in the graph. This results in tailored, context-aware suggestions, demonstrating how the combination of graph theory and advanced algorithms is a potent tool for harnessing information in complex datasets.

The rationale behind the usage of this technique:

(a) Effectively capture complex relationships in big data, ideal for modeling users, items, and content; (b) be scalable, handling large data volumes and easily adapting to new data types in big data scenarios; (c) offer personalized suggestions by understanding user preferences and behaviors; (d) address the sparsity in big data's user-item interactions better than traditional matrix approaches; (e) enhance recommendations by incorporating context like time, location, and social connections, and (f) reveal hidden patterns in data, improving recommendation quality and supporting collaborative filtering

The conditions for the optimal performance of this technique: (a) Graph construction is crucial for performance; nodes and edges must represent relevant data relationships, like user-product interactions in E-commerce; (b) Recommendation systems need to be accurate, personalized, and diverse to meet varied user preferences; (c) Sparse interaction data can be addressed using embedding learning to capture latent node and edge features; (d) Graph algorithms must be efficient and tailored to data characteristics and recommendation goals; and (e) Hybrid methods combining content-based and collaborative filtering help tackle the challenge of new users or items with minimal interaction history.

The limitations of this technique: (a) As the complexity and size of data grow, models require more computational resources. Scalability becomes a critical concern, especially for algorithms not designed to handle large-scale data efficiently; (b) Many models rely on rich interaction data to make accurate predictions. In scenarios where interaction data is sparse, such as with new users or items, the model's effectiveness can be significantly hindered. This is a common challenge in recommendation systems; (c) New items or users with limited historical data pose a "cold start" challenge. The model struggles to make accurate recommendations due to the lack of past interaction or behavioral data to learn from; and (d) Building and maintaining sophisticated models for varied and large datasets require substantial resources, including specialized expertise and computational power. This complexity often translates into higher costs and longer development times.

Table 7 evaluates and discusses research papers that have utilized graph-based models for big data in recommendation systems.

2.4 Rule-based models

The technique combines rule-based systems and graph algorithms to enhance the effectiveness of recommendation systems. Rule-based systems operate on a set of predefined rules or criteria, which are developed based on user behavior, item attributes, and other relevant data points. These rules can be segmented for different user groups based on demographics, past behavior, or other relevant factors, using data mining and analysis techniques to uncover patterns and trends that inform rule creation.

The technique synergizes graph algorithms like shortest path, clustering, or graph neural networks with rule-based systems to create a robust recommendation system. These algorithms analyze graph interconnections, uncovering patterns that enable highly personalized, context-aware recommendations based on user preferences.

This integration not only harnesses predefined rules based on user data and item attributes but also exploits the data graph's complex interconnections. Further enhanced by recommendation models like collaborative and content-based filtering, this approach ensures a comprehensive and effective strategy for recommendations.

The rationale behind the usage of this technique can be summarized as follows:

(a) The models are simpler and more transparent than complex machine learning models, making them ideal for sectors like finance or healthcare where explaining recommendations is crucial.

(b) They offer greater control and customization, as rules can be easily modified, making them suitable for dynamic environments.

(c) They efficiently manage large, varied datasets in big data contexts, quickly filtering relevant data using predefined criteria.

(d) They are stable and reliable, not relying on variable training data, and are apt for applications requiring consistent behavior over time.

(e) Rule-based models complement machine learning by handling well-understood, codifiable aspects, while machine learning deals with more complex data nuances.

The conditions for the optimal performance of this technique can be summarized as follows:

(a) Base the system's rules on sound logic and a deep understanding of user behavior and preferences.

(b) Use scalable and efficient algorithms and data

Table 7   Featuring research papers that have employed graph-based models for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Chen et al.[49] 2021	Amazon Musical Instruments	Fair	Fair	Acceptable	The authors presented TMER, utilizing a knowledge graph to model dynamic interactions between users and items over time. Simplifying the complexity of traditional neural networks, TMER captures user history and item relationships using attention mechanisms, enhancing both the accuracy and explainability of recommendations. The approach effectively integrates users' sequential behavior and broader contextual information, enhancing existing recommendation methods by emphasizing temporal and path-based analysis.
Qi et al.[50] 2022	PW dataset	Unsatisfactory	Unsatisfactory	Good	The authors introduced Web APIs Recommendation (WAR) to aid developers in discovering web APIs via keyword searches, streamlining app development. This method integrates planning, discovery, and API selection through data-driven techniques. Developers input keywords to find suitable web APIs. WAR uses a data graph with APIs as nodes and compatibility connections, creating subgraphs for each keyword set to outline app solutions.
Zhou et al.[51] 2019	MovieLens	Fair	Unsatisfactory	Fair	The authors developed a model merging cyber and social computing for Large-Scale Group Decision-Making (LSGDM) in online service social recommendation systems. They used a graph model in scholarly big data to depict LSGDM, profiling academic network decision-makers. This model assesses researchers' academic performance, connecting them through collaborative network interactions. Their method includes a network partitioning algorithm to identify key experts, and an improved random walk algorithm with restart.
Peng et al.[52] 2022	Comments dataset	Unsatisfactory	Unsatisfactory	Acceptable	The authors presented a privacy-focused stock recommendation method using federated learning, training word embeddings with encrypted investment forum data. Their federated meta embedding blends insights from private forums and public social media. they expanded this with federated graph meta embedding, adding graph-based modeling to integrate public and private.
Zhu[53] 2022	User-commodity behavior	Unsatisfactory	Unsatisfactory	Fair	The authors introduced a data fusion method for recommending mobile commerce services in tree-based networks. It entails storing tree-node relationships with redundant data to fully separate tree-type nodes for accessibility. The main objective is to efficiently store and retrieve structured data, using a tree synchronization model to boost server data buffering and access on mobile devices for enhanced efficiency.
Song et al.[54] 2023	Foursquare dataset	Fair	Unsatisfactory	Unsatisfactory	The authors created a precise spatiotemporal network-based recommender for suggesting Points of Interests (PoIs). Their model prioritizes spatial and temporal data, incorporating essential elements for accurate POI recommendations. It features well-designed meta-paths that account for both temporal and spatial aspects.

Table 7   Featuring research papers that have employed graph-based models for big data in recommendations.

(Continued)

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Wang et al.[55] 2023	Douban Movie	Unsatisfactory	Unsatisfactory	Unsatisfactory	The authors presented a framework, combining multi-community clustering and equitable decision fusion. This framework has three main parts: community exploration, local recommendations, and equitable decision fusion. It starts by creating overlapping communities through random walks on a user-item bipartite graph to connect users with similar motivations. The next step includes item ranking algorithms for each community.

structures to manage large data volumes without performance loss.

(c) Ensure the system can process data in near-real-time, crucial for adapting to rapidly changing user preferences.

(d) Personalize recommendations using individual user profiles and contextual information for relevance and effectiveness.

(e) Offer a diverse set of recommendations to prevent overfitting to known preferences and introduce users to new items.

(f) Design the model to handle sparse data, ensuring accurate recommendations in big data environments.

The limitations of this technique can be summarized as follows:

In big data contexts, rule-based systems face challenges with large data volumes, necessitating too many rules. More rules increase complexity and reduce agility, especially when adapting to new data patterns. These systems, limited by set rules, offer less personalization and are less adept at predicting preferences than machine learning models.

They tend to overfit historical data and struggle to adapt to new trends or unforeseen user behaviors. Their inflexibility in novel situations and data types, along with a focus on short-term, rule-defined patterns, means they may overlook longer-term user interests. They risk creating a "filter bubble", showing users only content like their past interactions.

Table 8 evaluates and discusses research papers that have utilized rule-based models for big data in recommendation systems.

3   Hybrid and Combined Approach

3.1   Ensemble models

This integrated approach leverages the ensemble technique, combining multiple recommendation algorithms or models to enhance the accuracy and reliability of recommendations. By merging various algorithms, it aims to utilize the strengths of each while addressing their individual weaknesses. A key aspect of this technique is diversity, employing different models like collaborative filtering, content-based filtering, and others in tandem. This ensures a more thorough analysis of data, capturing various dimensions of user preferences and behaviors.

The ensemble approach excels in handling big data, characterized by complexity and high dimensionality. It outperforms single-model systems by integrating various algorithms, including machine learning techniques like decision trees, neural networks, and k-nearest neighbors. This method employs strategies such as voting and averaging for different tasks, and advanced techniques like boosting and bagging to refine model weighting, significantly improving the precision and robustness of its predictions and recommendations.

The rationale behind this technique: These models in recommendation systems enhance accuracy by merging diverse algorithmic strengths, addressing user preferences and item characteristics. They reduce overfitting by balancing individual model biases, ensuring more stable predictions. These models adeptly manage big data's complexity and heterogeneity, including varied data types. Their diversity also fortifies system robustness, compensating for individual model weaknesses, particularly in handling data anomalies.

The conditions for the optimal performance of this technique: (a) Ensemble models improve by merging diverse models with different algorithms, architectures, and parameters, enhancing robustness in recommendations; (b) They need to balance

Table 8   Featuring research papers that have employed rule-based models for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Varlamis et al.[56] 2023	Electric power consumption	Unsatisfactory	Good	Good	The authors designed a collaborative filtering approach where users with comparable profiles share best practices. This system is bolstered by a sensor network that gathers extensive data on behavior and environment. By analyzing this data, the system identifies user habits and the specific micro-moments (or cues) that initiate these habits. It offers rule-based, personalized recommendations, tailoring these based on the user's latest activities and the information stored in the knowledge base.
Sumathi[57] and Akilandeswari 2020	Multilevel datasets	Fair	Unsatisfactory	Acceptable	The authors introduced an enhanced fuzzy weighted-iterative method to address the shortcomings in applying association rules focused on user requests and the visualization of rule discovery. This approach begins by integrating user feedback with post-processing to leverage semantic information. Following this, they developed rule schemas designed to accommodate and predict complex rules based on user expectations.
Zhang et al.[58] 2018	T10I4D	Unsatisfactory	Unsatisfactory	Good	The authors created MCRS, a course recommendation tool that enhances the Apriori algorithm using distributed computing. MCRS efficiently detects enrollment trends by initially preprocessing data with Hadoop, then analyzing it with Spark to extract association rules. These insights are then transferred to MySQL via Sqoop.
Liao and Yang[59] 2021	Author collected	Unsatisfactory	Fair	Fair	The authors designed a rule-based system that employs social network data and purchase behaviors to recommend fan pages. Their objectives encompassed user categorization, understanding the link between fan motivations and social network choices, analyzing fan page followers' online purchases, creating purchase-based profiles for enhancing e-business social network recommendations

complexity and simplicity to avoid overfitting or underfitting, ensuring effective pattern recognition; (c) In big data, these models must be scalable and efficient, using parallel processing and dimensionality reduction for large datasets; and (d) Regularization is crucial to prevent overfitting and enhance model generalization on new data.

The limitations are Ensemble models, combining multiple predictive models, require substantial processing power and memory, especially with big data. They risk overfitting, performing well on training data but poorly on new data, if individual models are too complex or poorly tuned. Integrating these models into existing systems can be complicated and resource-intensive.

Table 9 evaluates and discusses research papers that have utilized ensemble-based models for big data in recommendation systems.

3.2   Ranking models

The integrated approach to delivering personalized, relevant, and timely recommendations in big data environments involves a sophisticated interplay of data analysis, machine learning, and user-centric methodologies. It hinges on the use of advanced machine learning and data analytics techniques to interpret and navigate the immense datasets characteristic of these environments.

At the core of this process are sophisticated ranking algorithms that order items or services according to a user's preferences. These algorithms often utilize advanced methods like matrix factorization, deep learning, neural networks, and other pattern recognition techniques. The complexity of these algorithms can vary, ranging from simpler statistical models to more complex neural networks, each chosen based on the nature of the data and the specific requirements of the

Table 9 Featuring research papers that have employed ensemble-based models for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Jain et al.[60] 2022	Airline Dataset	Unsatisfactory	Fair	Fair	The authors analyzed customer reviews and ratings to predict product recommendations. They used a two-part approach: sentiment analysis of reviews using an LSTM model to assess sentiments about airline services, and evaluation of customer ratings for different aspects of these services. Combining these methods, they formed an ensemble to predict recommendations
Silva et al.[61] 2021	Iris, Glass, Sonar	Good	Unsatisfactory	Acceptable	The authors introduced methodologies for recommending classifier ensemble structures, using meta-learning for optimal settings of classifier type, number of base classifiers, and aggregation method. The aim is to efficiently establish a robust ensemble structure.
Hammou et al.[6] 2019	Yahoo! Webscope	Unsatisfactory	Good	Fair	The authors developed a distributed group recommendation system on Apache Spark for large-scale data. It merges dimension reduction with supervised and unsupervised learning, tackling the curse of dimensionality, identifying user groups, and improving prediction accuracy.
Huang et al.[62] 2019	MovieLens	Unsatisfactory	Fair	Fair	The authors introduced AMRE, which includes agents, a reward-function, and a roulette system. Each agent is a recommendation algorithm with a reward value based on accuracy, affecting its retention. The reward function modifies this value, rewarding accuracy and penalizing errors.

system.

In big data recommendation systems, machine learning models like decision trees, neural networks, and gradient boosting machines are essential. Continuous refinement through reinforcement learning, feedback loops, and A/B testing ensures their relevance and accuracy. Regular assessment with metrics like click-through and conversion rates helps adapt these systems to evolving user behaviors, maintaining their effectiveness in the dynamic big data landscape.

The rationale behind the usage of this technique can be summarized as follows: Traditional recommendation techniques struggle with the vast datasets typical in big data environments, but machine learning and AI-powered ranking models can process these efficiently, ensuring timely and relevant recommendations. These advanced models can detect complex, non-linear patterns in user behavior and item attributes, uncovering relationships that are too intricate for manual analysis or simpler algorithms. ranking models play a crucial role in personalizing user experiences by sorting items such as products, movies, and articles based on the user's past behavior and preferences. This enhances user satisfaction, as it allows users to find relevant content.

The conditions for the optimal performance of this technique:

(a) Model complexity: Balance is key; overly complex models risk overfitting and poor generalization, while too simple models may not capture data nuances.

(b) Algorithm selection and tuning: Essential for optimal performance, as different algorithms have varying strengths and weaknesses depending on the data and task.

(c) Handling data sparsity: Common in big data, effectively addressed using techniques like matrix factorization, embedding, or deep learning.

(d) Models must efficiently manage the vast volumes of big data, often necessitating distributed computing and efficient algorithms.

(e) Mitigating bias in recommendations is vital for ethical and effective system performance.

The limitations of this technique:

(a) Ranking models may unintentionally magnify existing biases in the training data, causing recommendations to be unfair or discriminatory, potentially favoring specific user groups or item types.

(b) There is a danger of models overfitting to complex, high-dimensional training data, impairing

their ability to generalize to new, unseen data and reducing real-world effectiveness.

(c) Focusing on accuracy, ranking models might compromise on diversity and novelty, often suggesting items too similar to users' past choices, thereby limiting exposure to new or varied options.

(d) Incorporating ranking models into existing systems and workflows in complex big data ecosystems can be difficult, demanding technical effort.

Table 10 evaluates and discusses research papers that have utilized ranking-based models for big data in recommendation systems.

4 Deep Learning

4.1 Context-aware models

The technique represents a significant evolution from traditional, generic recommendation methods to a more personalized and dynamic approach. It excels in leveraging contextual information such as location, time, user preferences, and environmental conditions, which enhances the relevance and personalization of recommendations. This method employs big data analysis, utilizing large and complex datasets to identify patterns, trends, and user behaviors. This improves the accuracy and effectiveness of recommendations by processing and analyzing diverse data types, from structured data like purchase histories to unstructured social media interactions.

This technique enhances recommendation systems by integrating detailed user profiling with advanced machine learning algorithms. These algorithms continuously evolve, adapting to new data and refining recommendations based on user interactions and current context. This approach not only improves the system's accuracy and relevance over time but also creates a more intuitive and responsive user experience. Moving from generic recommendations to a nuanced, context-aware strategy, it significantly

Table 10   Featuring research papers that have employed ranking-based models for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Huang et al.[63] 2019	GPS trajectory	Fair	Unsatisfactory	Good	The authors developed a recommendation system using the wide and deep model, which combines wide linear frameworks and deep neural networks for improved passenger identification in taxis. This system mimics experienced taxi drivers by focusing on passenger demand, road conditions, and potential earnings.
Li et al.[64] 2019	Wisdom Tourist	Unsatisfactory	Fair	Acceptable	The authors introduced a hybrid recommendation system that merges the Hierarchical Sampling Statistics (HSS) model with the MultiModal Visual Bayesian Personalized Ranking (MM-VBPR) algorithm. This system uses the DCA model to extract semantic correlations from image features and integrates them into the VBPR model, enhancing recommendation effectiveness.
Wang et al.[65] 2019	Video dataset	Good	Good	Acceptable	The authors proposed the Category-aided Multi-channel Bayesian Personalized Ranking (CMBPR) method. It combines content-based recommendations with a multi-channel BPR technique, using category-aided sampling and multi-channel sampling for a more nuanced understanding of user preferences in short video categories.
Lak et al.[66] 2022	Gowalla dataset	Fair	Unsatisfactory	Fair	The authors conducted a comparative study of various Bayesian personalized ranking algorithms, including NBPO and LightGCN, using multiple datasets. They focused on hyperparameter tuning to determine the most effective configurations for these recommendation algorithms.
He et al.[67] 2019	Amazon dataset	Acceptable	Fair	Unsatisfactory	The authors explored the integration of style features in personalized recommendations with their Style-aware BPR (SBPR) model. This approach uses a convolutional neural network to extract style features and a hierarchical gram matrix to capture stylistic elements, enhancing the representation of style in recommendation task.

elevates user satisfaction, marking a transformative development in recommendation system design.

The rationale behind this technique: (a) The technique enhances personalization by considering current context factors like location, time, device, and user mood, alongside past behavior; (b) It offers more accurate, real-time recommendations compared to traditional systems that rely on historical data, adapting to changing user preferences; (c) Suited for big data's volume, velocity, and variety, it efficiently processes and adapts to diverse and rapidly generated data; and (d) It increases user engagement by offering recommendations that align with the user's immediate needs and preferences,

The conditions for the optimal performance of this technique: (a) The ability to understand and utilize context effectively is crucial. This means the system should adapt its recommendations based on changing contexts, such as time of day, user location, or current events; (b) The model should be capable of learning individual user preferences and adapt recommendations accordingly. This requires sophisticated machine learning algorithms that can evolve with user behavior; and (c) The algorithms should be transparent and avoid biases. There should be mechanisms to avoid reinforcing stereotypes.

The limitations of the technique are: (a) Context modeling is complex due to the diversity in user location, time, social settings, and preferences. Efficiently capturing and processing this varied data poses a significant challenge; (b) Big data often has sparse datasets with many unobserved context-user-item interactions, making accurate predictions difficult; (c) Context-aware systems struggle with new users or items that lack interaction data, making accurate recommendations challenging; and (d) There is a risk of bias in models, favoring certain items or users. Ensuring fairness in recommendations is essential yet difficult.

Table 11 evaluates and discusses research papers that have utilized context-aware models for big data in recommendation systems.

4.2 Attention and memory networks

The integration of sophisticated concepts such as Attention Mechanism, Memory Networks, and Contextual Awareness significantly enhances the accuracy and relevance of recommendations in big data environments. Attention networks play a crucial role by prioritizing key user behaviors or items. For instance, if a user frequently purchases certain types of books, the system will focus more on these preferences when suggesting new books. This selective focus on aspects of data most relevant to a user's preferences ensures that the recommendations are more aligned with individual interests.

Memory networks store and retrieve user data, enhancing personalization in big data contexts by leveraging past behaviors for tailored recommendations. Contextual Awareness further refines this by considering factors like time and location, ensuring recommendations align with current user circumstances. The combination of these networks with attention networks, which focus on key data points, enables efficient navigation of extensive data, resulting in precise, individualized recommendations that reflect both comprehensive analysis and unique user needs.

The rationale behind this technique: Attention and memory networks are better suited for Big data environments as they can effectively process and learn from large datasets. Attention mechanisms in these networks focus on parts of the data more relevant to user preferences, leading to personalized and accurate recommendations. Memory networks enhance this by using a memory component to store and retrieve information from past user interactions, making recommendations more contextually relevant. These networks are particularly adept at handling sequential data, like in movie or music recommendations, understanding patterns over time. They also effectively manage data sparsity by focusing on the most informative interactions. Scalable, they adapt and learn efficiently with the growing volume of data in big data environments.

The conditions for the optimal performance of this technique: (a) The attention mechanism must differentiate essential from non-essential data points to boost recommendation precision; (b) The memory network must manage substantial data and retrieve pertinent information when necessary; (c) Scalability is crucial for coping with the growing amount, speed, and diversity of big data, requiring effective algorithms and expandable infrastructure; (d) Personalized recommendations based on individual user data are vital, necessitating an understanding of user behavior, preferences, and context; and (e) The system must

Table 11 Featuring research papers that have employed context-aware models for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Liu et al.[68] 2022	BEIJING dataset	Unsatisfactory	Good	Good	The authors developed Hydra, a multi-modal transportation planning system using deep learning. This system is sensitive to local points of interest and weather, integrating routing systems with urban data in a dual-layer framework. Hydra customizes routes for various travel modes, adapting to urban settings and user preferences.
Motwani et al.[69] 2023	Patient dataset	Fair	Fair	Good	The authors developed SPM for advanced patient monitoring. SPMR uses data from Ambient Assisted Living devices to monitor vital signs, symptoms, and activities, assessing and predicting patient health, and activating necessary services. It is compatible with technologies like IoT, ML, and AI, and integrates medical guidelines with patient data.
Gao et al.[70] 2020	Phoenix dataset	Unsatisfactory	Unsatisfactory	Fair	The authors developed N-PRA, a latent factor model-based algorithm, integrates social media contexts with user interests and dynamic POI popularity. It considers both POI types and geographic proximity, and factors in users' social networks, significantly improving the accuracy of personalized POI recommendations.
Chou et al.[71] 2020	CoMoDa, Frappe	Fair	Good	Acceptable	The authors developed DropTF, a fast tensor factorization method for context-aware recommendations using implicit feedback. DropTF excels by leveraging large volumes of unobserved data, showcasing the benefits of including all unobserved data in training a TF-based context-aware recommendation system.

continually learn from new data and adjust its recommendations.

The limitations are: (a) Attention and memory networks need considerable computational power, limiting their suitability for real-time environments; (b) Their effectiveness suffers from data scarcity, with accurate recommendations dependent on both the amount and quality of user data; (c) These networks are prone to overfitting on non-diverse, affecting their performance on new data; (4) They face challenges with new users or items due to their reliance on extensive historical data for accurate predictions; and (5) These models risk learning and propagating biases from their training data, potentially leading to unethical recommendations.

Table 12 evaluates and discusses research papers that have utilized attention and memory networks for big data in recommendation systems.

4.3 Neural graph-based

4.3.1 Graph Neural Networks (GNNs)

GNNs efficiently process data represented as graphs, comprising nodes (like users or products) and edges (indicating relationships). These networks excel in extracting features from graph structures by aggregating information from neighboring nodes, enabling them to understand complex node relationships and overall graph dynamics. Central to GNNs is the message passing mechanism, allowing nodes to update their information through interactions with adjacent nodes, revealing intricate data patterns. GNNs employ techniques like convolution and pooling to disseminate information across the graph. This aids in offering personalized, context-aware recommendations by considering user-item interactions and the user's network position, leading to precise predictions. GNNs also refine these recommendations by learning user and item characteristics.

The rationale behind this technique: GNNs excel in big data environments, adeptly modeling complex interactions between entities like users and products. They provide a deep understanding of data relationships. GNNs are particularly suited for graph-structured data common in big data scenarios such as social networks, enabling direct leveraging of this structure with less preprocessing. GNNs generate accurate results by considering data's relational context, identifying patterns that other methods might miss. They effectively tackle sparsity in big data.

The conditions for the optimal performance of this technique: (a) Given the vastness of big data,

Table 12   Research papers that have employed attention and memory networks for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Zhang et al.[72] 2022	Xing and Reddit	Fair	Unsatisfactory	Acceptable	The authors created A-PGNN, combining a Personalized Graph Neural Network (PGNN) and a dot-product attention mechanism. PGNN captures individual user behavior graph structures, focusing on the user's role in node embedding updates, unlike standard GNN models. The dot-product attention, inspired by Transformers, accurately represents the impact of past sessions on the current one.
Chiang et al.[73] 2023	MovieLens, Pinterest	Unsatisfactory	Fair	Good	The authors developed a recommendation system that dynamically adapts to user preferences by analyzing user behavior and recent interactions. It features a contextual item attention module for tracking preference changes and item relevance, and a multi-head attention module for handling diverse preferences. The system's accuracy is enhanced by including time-related item information, allowing for a personalized item representation in context.
Liu et al.[74] 2020	Patio/Lawn dataset	Fair	Fair	Good	The authors developed the AAMN model, which uses historical reviews and ratings with an attention mechanism to assess data relevance and create user and product profiles. The model excels in extracting crucial information from reviews and combines static and adaptive features through a non-linear fusion layer and a deep interaction layer for detailed insights into user-item interactions.
Yin and Feng[75] 2022	Amazon & Taobo	Good	Unsatisfactory	Acceptable	The authors developed a framework with enhanced attention, notably an external attention method, to better understand attention mechanism correlations and simplify complex self-attention processes. Aimed at visual tasks, it reduces computational load and captures relationships between samples. The framework also includes a multi-head mechanism for better performance.
Wang et al.[76] 2019	MovieLens	Unsatisfactory	Fair	Fair	The authors developed a recommendation system for top N items, using attention-based Seq2Seq and LSTM to analyze user behavior and optimize input selection and output simulation. It focuses on balancing training loss for better recommendation list generation and performance.

GNN should be scalable and efficient. Techniques like graph sampling, parallel processing, and efficient memory management are essential to handle large-scale graphs; (b) Choosing the right GNN architecture (like Graph Convolutional Networks (GCNs), Graph Attention Networks (GANs), etc.) and tuning parameters (like the number of layers, type of pooling, learning rate, etc.) is critical for optimal performance; (c) Regularization techniques (like dropout and weight decay) are important to prevent overfitting; and (d) The training strategy, including the choice of loss function, optimization, and training/validation split, impacts performance.

The limitations are: (a) GNNs struggle with scalability in big data scenarios, as they require processing a large number of nodes and edges, increasing computational complexity; (b) GNNs with multiple layers can cause over-smoothing, where node features become too similar, leading to information loss and decreased performance in recommendation tasks; (c) The presence of heterogeneous data (varied types of nodes and edges) in big data poses a significant challenge in developing GNN architectures that can effectively utilize this diversity; and (d) GNNs may inherit and amplify biases from training data, making ensuring fairness and avoiding discrimination in recommendations a major challenge.

Table 13 evaluates and discusses research papers that have utilized GNNs for big data in recommendation systems.

4.3.2 Neural Collaborative Filtering (NCF)
At its core, NCF uses collaborative filtering, predicting

Table 13 Research papers that have employed graph neural networks for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Cui et al.[77] 2024	MovieLens, LastFM	Unsatisfactory	Unsatisfactory	Acceptable	The authors proposed recommendation model combines graph diffusion with the Ebbinghaus curve to understand user interactions and memory patterns. The model identifies crucial memory paths in user interactions through a unique graph diffusion, facilitating a deeper analysis of these interactions. By integrating the Ebbinghaus curve, they personalize the model to reflect individual user's evolving interests and memory behaviors.
Wang et al.[78] 2023	MovieLens, Last.FM	Fair	Good	Unsatisfactory	The authors introduced MI-KGNN, a knowledge graph-based recommendation model that links users and items. It advances node representation through interaction analysis and a dual attention mechanism, allowing users and items to affect neighboring nodes' significance and highlight interaction patterns in the graph.
Sun et al.[25] 2023	MovieLens, Douban	Good	Unsatisfactory	Good	They authors proposed separated graph neural recommendation model which is a graph system that divides interaction networks into user and item segments. It effectively processes varied data, merges different methods for coefficient calculation, and uses a three-level attention mechanism for enhanced feature fusion, leading to effective/adaptable data propagation.
Sang et al.[79] 2023	YouTub, Yelp	Unsatisfactory	Unsatisfactory	Acceptable	The authors developed a model that uses a hierarchical heterogeneous graph neural network and adversarial training to improve user and item embeddings for recommendations. The model uses fake nodes to challenge its accuracy, adjusting to reduce reliance on extra regularization.
Xie et al.[80] 2022	DivMat	Unsatisfactory	Unsatisfactory	Good	The authors developed GraphDR, a neural network for recommendations, comprising a diversified preference network linking various nodes, a field-level heterogeneous graph attention network for learning node representations.

user preferences based on past interactions and similarities among users or items. NCF integrates deep learning to capture the non-linear and complex relationships in user-item interactions, going beyond traditional matrix factorization methods. The technique employs a multi-layer neural network architecture that allows for learning user-item interaction functions from data, offering a more flexible approach compared to traditional collaborative filtering.

The rationale behind this technique: (a) NCF overcomes sparsity and high dimensionality in big data, unlike traditional methods like matrix factorization, by leveraging deep neural networks to manage sparse data and extract features from high-dimensional spaces, thus enhancing recommendation accuracy; (b) NCF, with its neural network architecture, introduces non-linearity to effectively capture complex patterns in user-item interactions,

addressing the limitations of traditional collaborative filtering in understanding non-linear relationships; (c) NCF scales efficiently for big data using advanced optimization and neural network parallel processing; and (d) integrates extensive auxiliary data for enhanced user preference understanding.

The conditions for the optimal performance of this technique: (a) Choose neural network architecture that matches data complexity to prevent overfitting, (b) optimize hyperparameters like learning rate and neuron counts for better performance, (c) employ L2 regularization and dropout in complex models to generalize better, (d) apply advanced techniques such as embedding layers for complex interactions, and (e) Consider ethics and mitigate bias in recommendation systems for fairness.

The limitations are: (a) NCF models struggle with scalability in large datasets, leading to high

computational costs and processing times; (b) They face challenges with data sparsity and the cold start problem, making it hard to predict for new users/items; (c) Overfitting is common in NCF models, causing poor performance on unseen data; (d) These models lack transparency in decision-making, affecting user trust and satisfaction; (e) NCF models favor popular items, neglecting less popular ones and reducing recommendation diversity; (f) Their effectiveness is compromised in scenarios with limited historical data; and (g) Using deep learning in recommendation systems raises privacy concerns due to the need for detailed user data.

Table 14 evaluates and discusses research papers that have utilized NCF for big data in recommendation systems.

4.4 Autoencoders

Autoencoders, utilizing dimensionality reduction techniques, are highly effective in managing the complexity of big data. These techniques compress high-dimensional data into a more manageable form while preserving essential information. Autoencoders achieve this through a process that involves compressing data into a lower-dimensional space (encoding) and then reconstructing it back to its original high-dimensional space (decoding). This capability makes them particularly suitable for handling big data scenarios, like user-item interaction matrices in recommendation systems, which often contain sparse data.

Autoencoders, leveraging their neural network architecture, are adept in recommendation systems, capturing complex data relationships and generating detailed embeddings of users and items. These embeddings accurately reflect user preferences and item features, enhancing behavior and preference prediction. Through learning from user-item interactions, autoencoders improve collaborative filtering, identifying and utilizing patterns to recommend items matching user preferences. Their ability to record these patterns in a latent space is key for tailored recommendations, adapting to each user's unique behaviors and preferences. This capability makes autoencoders highly effective in personalized recommendation systems, notably enhancing the accuracy of predicting user preferences.

The rationale behind this technique are: (a) Autoencoders compress high-dimensional data into lower-dimensional, dense representations, ideal for recommendation systems with sparse data (many users and items, few interactions), facilitating the extraction of meaningful patterns; (b) They learn to encode data in a way that preserves relevant information, capturing underlying patterns in user preferences or item characteristics for accurate recommendations; (c) They excel in managing sparse data in recommendation systems, learning to reconstruct missing data or predict preferences with limited interaction data; (d) They adapt with variations like Variational Autoencoders and enhanced collaborative filtering; and (e) They capture non-linear relationships as neural networks, crucial for complex systems.

The conditions for the optimal performance of the technique: (a) The architecture of the autoencoder, such as deep or variational autoencoders, significantly affects performance. Tuning the number of layers and neurons per layer to match data complexity is necessary; (b) The choice of optimization algorithm (like Adam or Stochastic Gradient Descent (SGD)) and

Table 14   Research papers that have employed neural collaborative filtering for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Rehman et al.[81] 2023	MovieLens	Acceptable	Unsatisfactory	Fair	The authors created the weighted context-based neural collaborative filtering model, which improves neural collaborative filtering by using weighted contextual information. This model evaluates user interactions with items in different contexts, assigning varying weights to these contexts based on their relevance.
Wu et al.[82] 2023	Yelp	Fair	Unsatisfactory	Fair	The authors examined neural network-based recommendation models for accuracy, classifying them into collaborative filtering, content-based, and sequential types. They reviewed major studies and innovations in each group, provided insights, and discussed future research areas like basic concepts, model development, evaluation, and reproducibility in these systems.

learning rate is critical. A too high learning rate can miss the global minimum, while a too low rate may slow down convergence; (c) Optimizing batch size and the number of training epochs is important. Smaller batch sizes can improve generalization but may increase training time; (d) Autoencoders require substantial hyperparameter tuning due to their complexity and big data diversity for optimal performance.

The limitations of this technique: (a) Training autoencoders for recommendation systems is challenging due to sparse user-item interaction matrices, leading to difficulty in learning meaningful data representations; (b) Autoencoders face scalability problems with extremely large datasets, resulting in high training time and computational resource demands; (c) They have the risk of overfitting, especially with complex models and limited training data; (d) They struggle with providing recommendations for new users or items with minimal historical data, a significant issue in dynamic environments; (e) They are sensitive to data quality variations, affecting recommendation accuracy in big data environments.

Table 15 evaluates and discusses research papers that have utilized autoencoder for big data in recommendation systems.

4.5 Recurrent Neural Networks (RNN) and LSTM

RNNs excel at processing sequential data due to their ability to remember past inputs, which is crucial for pattern analysis and prediction. LSTMs, a specialized form of RNNs, are designed to learn long-term data dependencies. Their advanced architecture, featuring a

Table 15   Featuring research papers that have employed autoencoders for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Selvi and Kavitha[83] 2022	Medhelp dataset	Unsatisfactory	Acceptable	Good	The authors presented a secure health data recommendation method using a stacked de-noising convolution auto-encoder-decoder and a modified blowfish algorithm for privacy. This approach organizes patient data with Hadoop and analyzes both explicit and implicit patient information. It combines these features for effective recommendations.
Dong et al.[84] 2020	MovieLens, MovieTwe	Fair	Unsatisfactory	Fair	The authors integrated matrix factorization with autoencoder learning to analyze user/item ratings and learn low-dimensional representations, reducing time complexity. It includes user/item attributes to mitigate the cold start issue. The algorithm produces a prediction matrix by multiplying learned user and item features.
Nguyen and Cho[85] 2020	Gowalla dataset	Fair	Unsatisfactory	Acceptable	The authors created the Variational AutoEncoder Mixture Model (VAE-MM) to improve online behavior suggestions. It combines individual and group preferences using VAE. Individual preferences reflect past behaviors, while group preferences capture common interests. Enhancements like dropout layers and skip connections in the VAE increase its accuracy.
Drif et al.[86] 2020	Amazon Review	Unsatisfactory	Unsatisfactory	Acceptable	The authors developed EnsVAE for recommender systems, blending multiple sub-recommenders with a VAE to convert user-item interactions into interest probabilities, increasing accuracy. It offers enhanced data representation and adapts to various user interactions. EnsGG, a form of EnsVAE, combines GRU-MF CF with GloVe-content based filtering.
Zeng et al.[87] 2021	MovieLens	Unsatisfactory	Unsatisfactory	Fair	The authors created NCAR, an autoencoder framework for collaborative recommendations using implicit user trust. NCAR combines trust data and user-item interactions by embedding a user co-occurrence matrix and employing a neural recommendation process. It extracts this matrix from ratings, applies an autoencoder with correlation regularization for user embeddings.

cell state and three types of gates (input, output, and forget), allows them to manage information flow effectively.

In big data and recommendation systems, RNNs and LSTMs are highly effective. They handle large volumes of sequential user data, capturing temporal user behavior patterns for accurate, personalized recommendations. LSTMs are adept at understanding extended user behavior, crucial for distilling user preferences or item characteristics. Autoencoders complement RNNs and LSTMs by processing non-sequential data, like user profiles or item descriptions. They reduce high-dimensional data into simpler forms, facilitating easier handling by RNNs or LSTMs within a temporal framework.

The rationale behind this technique: (a) RNNs and LSTMs excel in handling sequential data like browsing and purchase histories, essential for predicting user actions in recommendation systems; (b) LSTMs remember long-term user interactions, crucial for understanding current interests and predicting future preferences in recommendation; (c) RNNs and LSTMs continuously update their understanding of user behavior, adapting to changing preferences, beneficial for evolving big data scenarios; (d) The architecture of RNNs and LSTMs enables them to identify intricate patterns, leading to more personalized recommendations; (e) RNNs and LSTMs are adept at processing unstructured data like text and images, vital for recommendations.

The technique's conditions for optimal performance: (a) The complexity of the model should be balanced. Overly complex models might overfit, while overly simple models might underperform; (b) Carefully choose the sequence length to capture relevant temporal dependencies; (c) Optimize these parameters for stable and efficient training; (d) Use methods like grid search or random search to find the optimal set of hyperparameters for your model; (e) Employ techniques to handle sparsity in recommendation data, such as using embedding layers; (f) Implement a mechanism for model adaptation in response to user feedback and evolving data patterns; and (g) integrating RNNs/LSTMs with techniques like CNNs for improved feature extraction.

The limitations of this technique: (a) RNNs and LSTMs are complex and computationally intensive, making scaling difficult for large-scale recommendation systems; (b) LSTMs struggle with very long sequences, affecting their ability to capture long-term user preferences; (c) RNNs and LSTMs perform poorly with sparse data and in cold start scenarios; (d) The sequential processing of RNNs and LSTMs may not efficiently adapt to the rapidly evolving data in recommendation systems; and (e) The opaque nature of RNNs and LSTMs makes it difficult to understand their decision-making.

Table 16 evaluates and discusses research papers that have utilized RNNs for big data in recommendation systems.

4.6 Sequence-aware models

The technique involves sequential pattern mining, where frequent patterns in user behavior are identified. This involves identifying and extracting patterns from sequences of user activities or behaviors, which helps in understanding and predicting future preferences. Temporal dynamics are crucial, as they consider the timing of user interactions and acknowledge that preferences evolve over time. Context-awareness is another key element, focusing on the context of user actions like location/time to enhance the recommendations.

Sequence-aware models often employ predictive algorithms like Markov chains, LSTM networks, or Gated Recurrent Unit (GRU) networks to forecast future user actions based on past sequences. By analyzing sequential data, these models offer highly personalized recommendations. For example, a music streaming service might suggest songs based on the sequence of previously listened tracks. Techniques like word2vec or Bidirectional Encoder Representations from Transformers (BERT) can be adapted to transform sequences of user interactions into vector space, making it easier to apply machine learning models.

The rationale behind this technique: (a) The models predict user preferences based on action sequences, offering nuanced insights in complex scenarios, such as E-commerce, to discern immediate interests and purchase intentions; (b) These models excel at capturing changes in user preferences over time, useful for platforms like streaming services to understand evolving tastes; (c) The models enhance accuracy and user experience by analyzing user interaction sequences for highly personalized recommendations; (d) These models use interaction

Table 16   Research papers that have employed RNNs for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Wang et al.[22] 2022	Beijing taxi dataset	Good	Fair	Unsatisfactory	The authors developed a method using a neural network with A* for improving personalized route recommendation. This model combines an attention-based RNN for calculating costs from the starting point to potential destinations and a position-aware graph attention network for estimating costs to the destination.
Zhang et al.[88] 2020	DBLP dataset	Unsatisfactory	Acceptable	Unsatisfactory	The authors introduced a system for academic social network friend recommendations using scholars' academic attributes from digital libraries. Their method employs an attributed random walk approach and a graph RNN framework, focusing on both network structure and scholar attributes.
Zhao et al.[89] 2016	Microblogging dataset	Acceptable	Good	Acceptable	The authors proposed using social and E-commerce network data for product recommendations. They utilized recurrent neural networks for E-commerce feature representation and modified gradient boosting trees for social network feature embedding, applying matrix factorization for cold-start product recommendations.
Shen et al.[90] 2020	Boston dataset	Fair	Unsatisfactory	Unsatisfactory	The authors explored the connection between listing descriptions and pricing, creating a price recommendation system for competitive pricing of new listings. Their approach used feedforward networks, LSTM, and mean shift algorithms, to enhance price prediction accuracy.

sequences in low user-item interaction settings to enhance recommendations, addressing data sparsity effectively; and (e) These models merge sequence data with context factors (time, location, and device) for more relevant recommendations in data-rich settings.

The technique's conditions for optimal performance: (a) Employing advanced sequence modeling techniques such as RNNs, Long LSTM networks, or Transformer-based models can significantly improve the ability to capture temporal dynamics in user-item interactions; (b) Personalization and context-awareness: The model should be capable of offering personalized recommendations by understanding individual user sequences. It should also consider contextual information like time of the day, location, or specific user states; (c) Maintain a balance between leveraging known user preferences and exploring new items to enhance user experience; and (d) Ensure the model is resilient to anomalies and generalizes well to new data or users through extensive validation and testing.

The limitations of this technique: (a) Models excel in short-term patterns but struggle with long-term user behavior due to diverse, prolonged interactions; (b) As data volume increases, these models become less efficient, demanding more computational resources and reducing scalability; (c) With limited historical data, especially for new users/items, models encounter the cold start problem, affecting recommendation accuracy; (d) Deep learning based sequence-aware models often overfit on large, diverse datasets, leading to poor performance on unseen data; (e) The opaque nature of deep learning recommendations complicates scenarios requiring clear reasoning; and (f) Sequence-aware models find integrating and processing varied data types (text, images, and videos) in big data challenging.

Table 17 evaluates and discusses research papers that have utilized sequence-aware models for big data in recommendation systems.

4.7   Convolutional operations

4.7.1   Graph Convolutional Networks (GCN)

GCNs for Big Data Recommendation seamlessly merge graph-based data representation with the powerful capabilities of convolutional neural networks, tailored to meet the demands of large-scale and diverse data sets. This innovative technique transforms the way data is handled, making it particularly effective for personalized and context-aware recommendations.

GCNs revolutionize recommendation systems by representing users and items as interconnected nodes in

Table 17 Featuring research papers that have employed sequence-aware models for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Chen et al.[91] 2023	MovieLens	Fair	Unsatisfactory	Unsatisfactory	The authors developed a data augmentation framework for sequential recommendation systems, employing counterfactual thinking in user behavior sequences. This framework includes seven models categorized into heuristic-based (four models), learning-based (two models), and reinforcement learning methods (one model).
Xu et al.[92] 2019	e-government	Fair	Acceptable	Good	The authors developed e-government recommendation system combining probabilistic semantic analysis with collaborative filtering, focused on user and item attributes, and utilizing historical data for group formation and sequence mining. It uses sequence mining tailored to e-government.
Cheng et al.[93] 2023	TripAdvisor	Acceptable	Fair	Unsatisfactory	The authors introduced Seq2CASE, a weakly supervised framework for extracting aspect scores from review comments. This method addresses the challenge of the absence of ground truth data for these scores, simplifying complex comments to estimate aspect scores. Seq2CASE demonstrates a strong correlation with actual user feedback.

a graph, capturing relationships through edges. This graph-based structure, combined with convolutional neural network operations, allows GCNs to process and learn from the network's topology effectively. GCNs excel in extracting rich feature representations from nodes, uncovering latent user preferences or item characteristics that raw data might not reveal. They aggregate information from neighboring nodes, enriching each node's context and enhancing recommendation quality. Also, GCNs incorporate both node and edge attributes, like user demographics or interaction types, making them highly effective and scalable for personalized recommendations in complex datasets.

The rationale behind this technique are: (a) GCNs excel in big data applications like social networks and E-commerce by modeling relationships in graph data; (b) They efficiently manage sparse data in recommendation systems by aggregating neighboring node information, aiding in predictions despite limited data; (c) GCNs enhance learning by incorporating additional data such as user demographics, crucial in big data for improved recommendations; (d) Capable of learning complex data representations; GCNs capture intricate user-item interactions for nuanced, personalized recommendations; and (e) By utilizing user-item graph structures, GCNs identify latent user behavior factors, boosting accuracy.

The technique's conditions for optimal performance: (a) The architecture of the GCN, including the number of layers, type of convolutional layers, and activation functions, needs to be carefully designed. Hyperparameter tuning, including learning rate, dropout rate, and regularization, is also essential for optimizing performance; (b) Employing graph sampling, mini-batch training, and efficient matrix operations are key strategies for managing large-scale data; (c) To prevent overfitting/ensure generalization, regularization techniques, dropout, and data augmentation methods should be utilized; and (d) Performance enhancement can be achieved by integrating GCNs with other models like matrix factorization, deep learning models (e.g., autoencoders), or other graph-based models.

The limitations of this technique are: (a) GCNs can struggle with scalability due to the computational complexity of convolutions over large graphs. Big data environments often involve very large user-item graphs; (b) GCNs might not perform optimally in sparse user-item interaction matrix as they rely on the graph structure for learning, and sparse data can lead to less informative graph structures; (c) Multiple graph convolutions can make node representations too similar, reducing accuracy; (d) GCNs have difficulty recommending for new users or items due to few connections; (e) GCNs, as deep learning models, lack

clear explainability; and (f) GCNs are prone to performance issues in noisy big data environments.

Table 18 evaluates and discusses research papers that have utilized GCNs for big data in recommendation systems.

4.7.2 Convolutional neural networks

The integration of CNNs in recommendation systems has transformed the way data is analyzed, especially in sectors like E-commerce and media streaming. The architecture of CNNs is intricately layered, enabling them to automatically detect and learn relevant features from raw data. This feature of learning is crucial for making accurate recommendations.

CNNs have a multi-layered structure where each layer detects different data aspects, from basic patterns in initial layers to complex features in deeper ones. Key to this are convolutional layers that filter and process data, and pooling layers that focus on significant features, with non-linear functions like ReLU enhancing pattern recognition. CNNs excel in analyzing visual content, extracting characteristics such as color and texture, and processing user interaction data (clicks, views, purchase history) for in-depth understanding of user preferences. They are often combined with other neural networks, like RNNs, to merge diverse data types, boosting recommendation accuracy and making CNNs vital in personalizing user experiences.

The rationale behind this technique: (a) CNNs excel in extracting complex features from high-dimensional data like images and videos, outperforming traditional algorithms in recommendations by identifying key patterns; (b) CNNs adeptly handle and interpret unstructured data such as images, text, and videos, making them invaluable in recommendation systems; (c) CNNs offer

Table 18   Research papers employed graph convolutional networks for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Yu et al.[5] 2022	LastFM, Douban	Acceptable	Unsatisfactory	Good	The authors introduced an adversarial framework with GCN. It features a GCN-based autoencoder for improving relation data by detecting intricate connections and reconstructing social profiles, clarifying each user's social relations. Also, they crafted a GCN-based social recommendation module sensitive to varying social relation strengths.
Yue et al.[12] 2023	Movielens, Douban	Fair	Unsatisfactory	Good	The authors developed the attribute-fusing graph convolutional network for recommendation systems, focusing on the role of attributes in representation learning. They employed an attention-based approach to combine user and item attributes into a unified representation within a <user, item, attribute> graph. A specialized Laplacian matrix is formulated for this graph.
Rang et al.[94] 2023	Yelp and Beibei	Unsatisfactory	Unsatisfactory	Fair	The authors developed the MBHCR model, which utilizes a heterogeneous graph to analyze complex user behaviors through varied interactions. This model features a multi-behavior relational aggregator to identify common interaction patterns, addressing user sparsity issues. Additionally, it incorporates a behavior comparison learning enhancer to differentiate and highlight key aspects of primary and secondary user behaviors.
Liu et al.[95] 2023	Gowalla & Yelp	Fair	Unsatisfactory	Good	The authors presented a method for modeling multi-grained popularity features and high-order connectivity, focusing on diverse user preferences in popularity features. They created the JMP-GCF model, which utilizes popularity-aware embeddings and a joint learning strategy to analyze multi-level popularity signals.
Yang et al.[96] 2023	Amazon Musical	Good	Unsatisfactory	Fair	The authors developed the PGIR model that fuses enhanced graph convolution with review properties, aligning property data with review text for in-depth content analysis. This model focuses on complex user-item relationships and employs sophisticated graph convolution to capture dynamic user-item features and collaborative patterns.

advanced understanding of context and spatial relationships in data; (d) CNNs' adaptability through transfer learning allows systems to use pre-trained models, reducing the need for new data; (e) CNNs boost collaborative filtering by leveraging auxiliary data to address cold start and data sparsity, thus enhancing recommendation accuracy; and (f) CNNs excel in recommending visual content, like fashion or movies, due to their proficiency in image and video analysis, ensuring relevancy.

The technique's conditions for optimal performance: (a) Selecting the appropriate number and types of layers (convolutional, pooling, and fully connected), filter sizes, and network depth is critical. Integrating CNNs with other neural networks like RNNs enhance the understanding of sequential patterns in user behavior, especially for recommendation systems; (b) Optimizing parameters such as learning rate, batch size, and epochs is crucial for CNN performance. Advanced techniques like Bayesian Optimization can further refine this process; (c) Regularization methods like dropout, L1/L2 regularization, and data augmentation are vital in big data scenarios to avoid overfitting; and (d) Utilizing pre-trained models is advantageous, especially with limited labeled data.

The limitations of this technique: (a) CNNs are not ideal for processing non-visual, non-spatial data like user behavior or preferences in recommendation systems; (b) Due to their deep architectures, CNNs risk overfitting, especially with insufficient training data, which is often the case in specific recommendation tasks; (c) CNNs require substantial computational resources, making them impractical for organizations with limited hardware or in scenarios demanding quick updates; (d) CNNs face challenges in scaling for large datasets and lack flexibility due to their fixed architecture; (e) The "black box" nature of CNNs is a drawback in where understanding the rationale behind recommendations is crucial; and (f) CNNs struggle to make accurate recommendations for new users or items with minimal historical data.

Table 19 evaluates and discusses research papers that have utilized CNNs for big data in recommendation systems.

4.8 Deep Reinforcement Learning (DRL)

DRL algorithms in big data recommendation systems learn through trial-and-error by interacting with vast datasets of user preferences and behaviors. Central to these systems is an agent, usually an algorithm, that makes decisions based on the observed user data, aiming to maximize rewards. DRL adapts its recommendations continuously, enhancing personalization by understanding and predicting user preferences. It excels in sequential decision-making, focusing on long-term rewards and balancing the exploration of new recommendations with the exploitation of known preferences. This approach is vital in dynamic big data environments, allowing DRL to effectively adapt to constantly evolving user preferences and data, ensuring more relevant and engaging recommendations over time.

The rationale behind this technique: (a) DRL thrives in changing environments, such as recommendation systems where user preferences and trends are always shifting. It continuously updates its learning based on new data, ensuring recommendations stay relevant; (b) DRL excels at identifying complex patterns, offering personalized recommendations for each user, even in large-scale scenarios; (c) DRL focuses on long-term user engagement, offering recommendations that enhance satisfaction and retention; and (d) DRL effectively manages recommendation systems by strategically assessing the long-term impact of decision sequences.

The technique's conditions for optimal performance: (a) In DRL for recommendation systems, the reward function must focus on long-term user satisfaction and engagement, incorporating complex metrics beyond immediate clicks or views; (b) DRL algorithms need to be scalable for big data and stable for consistent learning. Techniques like experience replay, target networks, and distributed learning are crucial; (c) DRL models in recommendation systems should continuously learn and adapt to changing user preferences in dynamic environments, requiring online learning; and (d) DRL must strike a balance between exploring new recommendations and exploiting known user preferences, using strategies like ε-greedy or upper confidence bound.

The limitations of this technique are: (a) DRL models are complex and need substantial computing resources, making them less suitable for real-time systems that require quick responses; (b) DRL faces challenges in recommendation systems with new items

Table 19   Research papers that employed CNNs for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Liu et al.[21] 2021	Amazon Product Review	Good	Unsatisfactory	Acceptable	The authors created the NCTR model, a hybrid neural network that combines text and rating data for item recommendation. NCTR uses a convolutional neural network to analyze text context and a fusion layer to integrate these features. Also, it employs multilayer perceptrons to manage the nonlinear interactions of combined item and user latent features, enhancing the accuracy.
Chen et al.[97] 2022	Douban and Yelp	Acceptable	Unsatisfactory	Fair	The authors created a recommendation system that combines a heterogeneous information network and deep learning. This system uses network embedding and deep learning to analyze auxiliary data and extract features from reviews. It employs an attention mechanism and dual parallel CNNs for processing user and item word vectors.
Ahammad et al.[98] 2023	Yelp	Good	Acceptable	Unsatisfactory	The authors developed a hotel recommendation system combining user collaboration, preference analysis, and similarity-based recommendations. It utilizes collaborative filtering and classified data management, incorporating capsules with convolutional kernels for in-depth feature analysis.
Wang et al.[99] 2023	Amazon & Yelp	Fair	Unsatisfactory	Fair	The authors introduced the AUR framework, combining an uncertainty estimator with traditional recommendation models. Using aleatoric uncertainty, it aims to improve niche item recommendations while preserving general performance. AUR is successfully integrated with matrix factorization, LightGCN, and VAE models.

or users, as it depends heavily on extensive interaction data; (c) Developing effective reward functions for DRL is difficult, particularly in recommendation systems where rewards are indirect or delayed; (d) In DRL, finding a balance between introducing new recommendations and adhering to known user preferences is challenging. Too much exploration can irritate users, while too little can lead to suboptimal strategies; and (e) DRL models may perpetuate and intensify biases present in their training data, leading to fairness issues in recommendations.

Table 20 evaluates and discusses research papers that have utilized DRL for big data in recommendation systems.

4.9   Self-attention mechanisms

The self-attention mechanism in models is pivotal in understanding and processing user interaction data, such as clicks, views, and purchases. By focusing on different parts of the input sequence, the mechanism is adept at capturing the dependencies and relationships between various items in a dataset. This is particularly useful as it enables the model to not just consider individual items in isolation but to understand how one item in a user's history might influence their interest in another. Consequently, this approach allows the model to dynamically adapt to the input data, emphasizing the most relevant parts for making predictions or recommendations.

Self-attention enhances user experience in E-commerce, content streaming, and social media by accurately discerning user history for personalized recommendations. It contextually analyzes behavior, distinguishing between casual browsing and specific purchase intentions, and dynamically adjusts to align recommendations with current preferences and future behavior.

The rationale behind this technique are: (a) Self-attention is adept at handling sequential data like user histories in recommendation systems. It processes sequences in parallel, unlike traditional models like RNNs or LSTMs, enhancing training and inference efficiency; (b) Self-attention is effective in recognizing relationships between distantly placed items in a user's interaction history, essential for recommendation systems. It computes attention across the sequence, allowing full contextual consideration; (c) Inherent flexibility and scalability make self-attention ideal for

Table 20 Research papers that have employed DRL for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Huang et al.[13] 2023	Geolife, Chengdu	Fair	Fair	Unsatisfactory	The authors created FedDSR, a daily schedule recommendation system using deep reinforcement learning within a federated learning framework. They applied curriculum learning for better local optimization and generalization and introduced a similarity aggregation algorithm to refine model quality using locally uploaded parameters. The system operates on encrypted local devices, ensuring privacy.
Zhou et al.[100] 2021	DBLP, ResearchGate	Fair	Unsatisfactory	Acceptable	The authors created a Hierarchical Hybrid Network (HHN) using deep reinforcement learning and a modified random walk algorithm for analyzing large dataset correlations in recommendation systems. This system, tailored for collaborative scholarly big data projects, effectively evaluates complex multi-level entity interactions, showcasing the HHN's capability in intelligent routing and data analysis.
Wang et al.[101] 2023	New York Check-in	Acceptable	Fair	Acceptable	The authors created a deep interactive reinforcement learning framework for studying geo-human interactions, consisting of representation and imitation modules. The representation module converts geo-human interactions into embeddings (state), and the imitation module, acting as a reinforced agent, recommends the next POI based on these states, imitating user visit patterns. The model's success improves through feedback.

big data environments with high volume, variety, and velocity. It adapts well to large datasets and complex user behavior patterns; and (d) It improves recommendation systems' accuracy by deeply understanding user preferences and behavior, leading to precise, personalized recommendations.

The technique's conditions for optimal performance: (a) Self-attention mechanisms must balance user preferences over time with immediate interests, necessitating an adept attention design; (b) These models, complex in nature, risk overfitting, especially with large datasets. Employing regularization methods like dropout and early stopping is crucial; (c) The success of self-attention depends on hyperparameters like attention head count, embedding size, and learning rates. Precise tuning of these is key; (d) In big data contexts, adapting self-attention to user behavior changes and contextual shifts, possibly by adding temporal or contextual elements, is vital; and (e) While self-attention is effective, optimal performance often requires integration with other elements, such as CNNs for feature extraction or RNNs for sequence processing, customized to the recommendation system.

The limitations of this technique are: (a) Self-attention requires significant memory resources, as it involves storing and processing all pairwise interactions between elements in the input sequence. In big data contexts, where the input size can be extremely large, this can lead to excessive memory demands; (b) While self-attention is adept at capturing global dependencies, it may not always efficiently capture very long-term dependencies in sequences, which is important in recommendation systems where historical data plays a crucial role; (c) In recommendation systems, self-attention can underperform due to sparse data and cold start issues; (d) Self-attention risks overfitting in data-rich but pattern-repetitive environments; (e) Self-attention can amplify biases, reducing recommendation diversity and perpetuating existing patterns; and (f) Self-attention models are computationally intensive, increasing training time.

Table 21 evaluates and discusses research papers that have utilized self-attention mechanisms for big data in recommendation systems.

5 Algorithmic and Mathematical Modelling Method

5.1 Probabilistic and statistical models

The approach to big data challenges employs machine

Table 21   Research papers that have employed self-attention mechanisms for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Meng et al.[102] 2023	MIND-small and Adressa-1week	Acceptable	Fair	Acceptable	The authors introduced GAINRec, a news recommendation model that leverages both personalized preferences and collective behavior patterns. It uses self-attention and cross-attention mechanisms to understand user preferences from news titles semantically, and a global transition graph to capture common behavior patterns. GAINRec combines these insights with a heterogeneous transition graph attention network, improving news recommendation accuracy and trustworthiness.
Sun et al.[103] 2019	Yoochoose and Diginetica	Fair	Unsatisfactory	Good	The authors developed SANSR, a self-attention network for session-based recommendation. It uses attention mechanisms for parallel processing, identifying relevant items in a session and assigning weights for predicting the next item. It includes an embedding layer, self-attention blocks, and a feedforward network, concluding with incremental training in the prediction layer.
Wang et al.[104] 2019	CAMRa2011 & Meetup	Good	Fair	Fair	The authors introduced SACML, combining neural networks with metric learning for recommendations. It uses self-attention to determine the significance of group member interactions, accumulating group preferences and applying collaborative metric learning for group recommendation tasks.
Yang et al.[105] 2023	MIND news recommendation	Unsatisfactory	Unsatisfactory	Good	The authors proposed the MIAR method, using multichannel information fusion for user interest recommendations. It combines user-clicked news and title embeddings with candidate news. MIAR consists of an interactive framework (MIF) and a distributed framework with an interest activation module, refining user profiles and personalizing user representations.

learning algorithms, notably those based on Bayesian methods, to provide accurate and personalized recommendations to users. Central to this approach is the use of statistical inference, which infers user preferences and patterns from data, embracing the complexities and uncertainties inherent in big data. It frequently uses probabilistic methods, like Bayesian networks, probabilistic latent semantic analysis, and Markov decision processes, to manage the uncertainty in predicting user preferences. These methods estimate the likelihood of a user's preference for an item, based on their history and the item's characteristics.

In this approach, Statistical Learning methods like CF and SVD analyze user behavior and preferences to predict interests. Additionally, advanced models incorporate contextual factors like time and location, refining recommendations. These algorithms adapt and improve over time, enhancing prediction accuracy in complex data environments.

The rationale behind this technique: (a) Probabilistic models are adept at handling the uncertainty and complexity inherent in user behavior and preferences. Big data environments often involve noisy, incomplete, or inconsistent data, and probabilistic models can effectively manage these challenges by making probabilistic inferences about user preferences; (b) Statistical models allow for the personalization of recommendations. They can analyze large datasets to identify patterns and trends specific to an individual. This enables recommendation systems to offer more tailored suggestions, improving user satisfaction; (c) These models are adaptable to various data types like ratings and clicks, crucial in big data's diverse data landscape; (d) Leveraging historical data and user interactions, these models precisely predict preferences, leading to efficient, learning-based recommendations; (e) Proper regularization techniques enable these models to avoid overfitting in high-

dimensional big data, maintaining generalizable and relevant recommendations; and (f) Probabilistic and statistical models integrate contextual details (time, location, etc.) into recommendations, enhancing their relevance and timeliness.

The technique's conditions for optimal performance: (a) The complexity of the model should be balanced. Complex models may overfit the training data and not generalize well to new data. Regularization techniques and cross-validation can help in managing this balance; (b) These models should account for changes over time in user preferences and item popularity. Incorporating temporal dynamics can improve recommendation quality; (c) Big data environments require models that can efficiently scale with data volume, using techniques like stochastic gradient descent, online learning, and distributed computing; (d) Probabilistic models, such as matrix factorization techniques, are effective in dealing with sparse datasets, where many user-item interactions are unknown; (e) Models need to personalize recommendations based on user profiles and contextual information, adapting to individual preferences and contexts; and (f) Addressing the challenge of making recommendations for new users or items with limited data through content-based filtering or hybrid models.

The limitations of this technique: (a) These models are prone to overfitting, especially when the data is sparse but high in dimensionality (common in recommendation systems). Overfitting leads to models that perform well on training data but poorly on unseen data; (b) Recommendation systems often suffer from data sparsity issues. Probabilistic and statistical models might not perform well under these conditions as they rely on sufficient data to understand and predict user preferences accurately; (c) These models have difficulty dealing with new users or items (cold start problem) because they lack historical data to make accurate recommendations; (d) Input data biases can be amplified in the recommendations, posing issues in scenarios requiring fairness and diversity; (e) Selecting and tuning the right model is complex and resource-intensive, necessitating extensive experimentation; and (6) Often designed for short-term preferences, these models may fail to capture and predict long-term user behavior.

Table 22 evaluates and discusses research papers that have utilized probabilistic/statistical for big data in recommendation systems.

5.2 Clustering-based models

The approach to data segmentation and recommendation involves employing clustering algorithms, which are essential to the core of these models. These algorithms, including K-means, hierarchical clustering, and density-based spatial clustering of applications with noise, are selected based on the specific requirements and characteristics of the data. Their function is to divide the data into groups or clusters based on similarities, where each cluster represents a group of users or items sharing similar characteristics. For instance, one cluster might consist of users who prefer action movies, while another could include those who frequently purchase sports equipment. Feature selection is vital in this process, focusing on the most relevant data features for accurate, efficient clustering. The model then uses these clusters to generate tailored recommendations, considering both the cluster's common attributes and individual user preferences within that group.

The rationale behind this technique: (a) Clustering algorithms, such as K-means, hierarchical clustering, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN), can efficiently handle large datasets. They group similar items or users into clusters, reducing the complexity of the data. This makes it easier to process vast amounts; (b) By clustering similar items or users, these models can offer more personalized recommendations. For instance, a user can be recommended items that are popular within their cluster but might not be widely known outside of it. This approach is useful when niche items get recommended alongside popular ones; and (c) Clustering helps in uncovering hidden patterns and relationships in the data, leading to more accurate recommendations.

The technique's conditions for optimal performance: (a) The choice of clustering algorithm greatly impacts performance. Algorithms like K-Means, Hierarchical Clustering, or DBSCAN should be chosen based on the dataset's characteristics (size, density, and dimensionality). The algorithm should be efficient and scalable to handle big data volumes; (b) Finding the optimal number of clusters is crucial, using methods like the elbow method, silhouette analysis, or gap statistics to avoid overfitting and underfitting; and (c) In big data, clustering models need to be dynamic

Table 22 Featuring research papers that have employed probabilistic/statistical for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Cui et al.[106] 2023	Amazon Beauty & LastFM	Acceptable	Acceptable	Good	The authors presented the Factors Mixed Hawkes Process (FMHP), an approach for incremental recommendation systems that analyzes key factors influencing event generation. FMHP transforms events into sequences driven by these factors, using the Hawkes process for modeling based on past events. It integrates various functions to evaluate different event intensities and employs an incremental updating algorithm, enabling real-time adjustments in event intensity, thus improving the system's adaptability to changing user interactions.
Liu and Zhao[24] 2023	Amazon food dataset	Acceptable	Fair	Fair	The authors created SAMF, a recommendation system combining sentiment analysis, matrix factorization, topic modeling via latent Dirichlet allocation, and deep learning with BERT. They generate topic distributions from reviews to shape user and item matrices, forming a comprehensive user-item preference matrix for rating predictions.
Mei et al.[107] 2020	AIS dataset	Fair	Fair	Acceptable	The authors devised an innovative algorithm for recommending substitute ports for container ships. This method, distinct from standard adjacency matrix techniques, utilizes NLP to analyze AIS data. It creates sequential berth records and models ports as vectors in a spatial embedding. The algorithm proposes ports comparable to the originally planned but inaccessible ones and verifies these suggestions by examining sailing distances.
Zhu and Sun[108] 2023	Author collected	Unsatisfactory	Acceptable	Fair	The authors studied a specialized educational information system that uses algorithms to recommend personalized content. The system analyzes user history and tag attributes, incorporating data pre-cleaning for accuracy. It uses a clustering algorithm for more efficient and in-depth analysis. The study also applies collaborative filtering, enhanced with information entropy and standard deviation, to identify user similarities.
Xia et al.[109] 2023	Taxi trajectory	Unsatisfactory	Acceptable	Good	The authors developed BiA*-ACO, a hybrid algorithm that efficiently identifies taxi routes in urban networks. It merges the bidirectional A-star algorithm's cost estimation with the Ant Colony algorithm's heuristic, enhancing search efficiency. Additionally, it improves the ant colony algorithm's pheromone update rules by incorporating the best routes from each iteration.

and adaptable for rapid preference changes, enabling periodic cluster updates with low overhead.

The limitations of this technique are: (a) While clustering is effective for grouping similar items or users, it may not scale well with extremely large datasets in big data scenarios. As the number of users and items increases, computational complexity can become a significant challenge; (b) Clustering tends to group similar items or users together, which can lead to a lack of diversity in the recommendations. This homogeneity might not always align with the users' desire for varied recommendations; and (c) The

effectiveness of clustering-based recommendations relies heavily on the quality of clusters. Poorly defined clusters result in inaccurate or irrelevant suggestions.

Table 23 evaluates and discusses research papers that have utilized clustering-based models for big data in recommendation systems.

6 Experimental Evaluation

6.1 Evaluation methodology

The following methodology is utilized in conducting the experimental evaluations:

Table 23   Research papers that have employed clustering-based models for big data in recommendations.

Paper/year	Dataset	Interpretability	Scalability	Efficiency	Description
Rostami[110] 2022	Author collected	Unsatisfactory	Fair	Good	The authors developed a two-phase hybrid food recommender system to overcome limitations like disregarding ingredients, time factors, and new users/items. The first phase involves graph clustering for content-based recommendations, and the second uses deep learning for user and food item clustering. The system also incorporates strategies to handle time-related and community factors.
Nie et al.[1] 2023	Movielens	Fair	Fair	Acceptable	The authors created the UCMF model for better cross-domain recommendations. It uses graph neural networks to fuse user data from multiple domains, improving user representation, especially in data-scarce situations.
Yue et al.[12] 2023	MovieLens and Douban	Acceptable	Fair	Good	The authors proposed TCCF, a collaborative filtering algorithm, combines Time Correlation Coefficient (TCC) and optimized CSK-means clustering. It simplifies large data problems through clustering, grouping similar users for efficient recommendations. This includes an improved K-means algorithm using Cuckoo search optimization and a time factor to track evolving user interests. PTCCF further personalizes this approach, focusing on user preference patterns.

• **Evaluating individual techniques:** After a comprehensive review of papers presenting algorithms employing a specific technique, we identify the paper with the greatest impact. The algorithm detailed in this influential paper is chosen as the representative for its respective technique. To determine the most significant paper among those reporting algorithms using the same technique, we consider various factors including its innovative contributions and recency.

• **Ranking the techniques within the same sub-category:** Initially, the Mean Average Precision (MAP) for each algorithm is calculated. Next, the MAPs of the algorithms using the same technique are averaged to determine the MAP representative of that technique. Finally, within each primary sub-category, the techniques are ranked based on their computed average MAPs.

• **Ranking the sub-categories within the same category:** First, the MAPs of the algorithms using the same sub-category are averaged to determine the MAP representative of that sub-category. Then, within each primary category, the sub-categories are ranked based on their computed average MAPs.

• **Ranking the categories within the same method:** First, the MAPs of the algorithms using the same category are averaged to determine the MAP representative of that category. Then, within each primary method, the categories are ranked based on their computed average MAPs.

6.2   Evaluation setup

In the process of tuning the hyper-parameters for the selected algorithms, our initial step involves experimenting with various parameter values for each algorithm. Subsequently, we select the set of parameter values that enabled each algorithm to attain its optimal recommendation performance. We list below these values:

• Yu et al.[5]: Number of layers: 3; activation function: ReLU; dropout rate: 0.3; learning rate: 0.001; weight decay: 0.0005; number of epochs: 200; early stopping: 10 epochs without improvement; neighborhood sampling: 10.

• Wang et al.[22]: Number of layers: 2; number of units per layer: 128; dropout rate: 0.5; learning rate: 0.001; batch size: 32; number of epochs: 20; sequence length: 50; and embedding dimension: 100.

• Hu et al.[27]: min_df: 0.01 (exclude terms that have a document frequency lower than 1%); max_df: 0.8 (exclude terms that have a document frequency higher than 80%); ngram_range: (1, 2) (include both unigrams and bigrams); max_features: 10 000; λ: 0.01; learning

rate: 0.001; number of epochs: 50; and batch size: 32.

• Yan et al.[32]: Neighborhood size: 20; rating scale: 1 to 5; minimum number of ratings: minimum 5 ratings per item; data sparsity threshold: at least 10 interactions per user; implementation specific parameters: batch size = 32, number of epochs = 10: and learning rate = 0.01.

• Yi et al.[35]: Number of factors: 50; regularization parameter for users: 0.1; regularization parameter for items: 0.1; learning rate: 0.02; and maximum iterations: 500.

• Guan et al.[39]: Tolerance: 0.0001; random seed: 42; maximum iterations: 1000; and learning rate = 0.01.

• Ioannidis et al.[42]: rank: 10; regularization parameters ($\lambda\_users = 0.1$; $\lambda\_items = 0.1$; and $\lambda\_time = 0.1$); learning rate: 0.01; maximum iterations: 1000; initialization method: random; convergence criterion: 0.001; and loss function: RMSE.

• Qi et al.[46]: Number of neighbors: 20; threshold for similarity: 0.5; and minimum number of ratings: 5.

• Zhang et al.[72]: Embedding size: 128; attention layers: 2; attention heads: 4; key size: 32; value size: 32; learning rate: 0.001; batch size: 128; number of epochs: 20; regularization term: 0.0001; and dropout rate: 0.5.

• Wang et al.[78]: Number of layers: 3; number of hidden units: 64; embedding size: 128; learning rate: 0.001; number of epochs: 30; batch size: 100; dropout rate: 0.5; regularization factor: 0.0001; activation function: ReLU; edge dropout rate: 0.1.

• Selvi and Kavitha[83]: Number of layers: 3; number of neurons in each layer: (input layer: 100 neurons; hidden layer: 50 neurons; and output layer: 100 neurons); activation function: ReLU; learning rate: 0.001; batch size: 128; number of epochs: 100; and dropout rate: 0.5.

• Sun et al.[103]: Number of layers 6: model dimension: 512; number of heads in multi-head attention: 8; inner layer dimension of feed-forward networks: 2048; dropout rate: 0.1; positional encoding dimension: 512; learning rate: 0.0001; and batch size: 64.

• Rostami et al.[110]: Number of clusters: 5; initialization method: k-means++; maximum iterations: 300; convergence tolerance: 0.0001; random state: 42; N_init: 10.

We search for publicly available codes for these algorithms. We could obtain codes for the following

Refs. [21]※, [22]†, [27]‡, [42]§, [50]¶, and [72]☆. For the remaining representative papers, we develop our own implementations using TensorFlow, as described by Sinaga and Yang[111]. We train these implementations using the Adam optimizer, as suggested by Sinaga and Yang[111]. TensorFlow's APIs provide users with the flexibility to create their own algorithms[112]. Our development language is Python 3.6, and we utilize TensorFlow 2.10.0 as the backend for the models.

6.3   Compiling datasets for evaluations

• MovieLens dataset©: It is a repository of movie ratings. It integrates a vast array of user ratings along with movie metadata like titles, genres, release years, and user demographics. The 20M version, which we focus on, is substantial, featuring 2.0×10^7 ratings and 4.65×10^5 tag applications for 2.7×10^4 movies by 1.38×10^5 users.

• Amazon Electronics dataset¤: It offers user reviews and detailed product information from Amazon. It encompasses an extensive array of electronic product reviews, featuring user ratings, textual feedback, and helpfulness votes. It includes product metadata descriptions, categories, pricing, brands, and image features. It encapsulates around 3.5×10^7 reviews up to March 2013.

6.4   Experimental results

We run the algorithms using an Intel(R) Core(TM) i7-6820HQ processor at 2.70 GHz, with 16 GB RAM, operating on Windows 10 Pro. We utilize the Mean Average Precision (MAP) metric for the evaluations. MAP calculates the average precision across different ranks, emphasizing the order of retrieved items. Precision at K (P@K) gauges the ratio of relevant items in the top K results.

• Table 24 shows the overall average MAP scores of the selected algorithms, it also shows the following: (1) the ranking the techniques within a same sub-category; (2) the ranking the techniques within a same sub-category; (3) the ranking the categories within a same

※ https://github.com/bioannidis/Coupled_tensors_graphs
† https://github.com/qlyseven/source-code
‡ GitHub - luojia527/NCTR_master: python
§ https://github.com/CRIPAC-DIG/A-PGNN
¶ https://github.com/bigscity/NASR
☆ https://github.com/facebookresearch/fastText
© https://grouplens.org/datasets/movielens/
¤ https://www.kaggle.com/datasets/saurabhbagchi/amazon-electronics-data

Table 24 **Overall average MAP of each selected algorithm using MovieLense dataset (ML) and Amazon Electronics dataset (AE).**

Method	Category	Sub-category	Technique	Paper	Dataset	Average MAP	Technology rank	Sub-category rank	Category rank	Method rank
User and Item similarity based methods	Content-based filtering	–	–	[27]	ML	0.67	–	–	3	3
					AE	0.65				
	CF	Item-based CF	–	[32]	ML	0.73	–	2	1	
					AE	0.69				
		MF	Probabilistic MF	[35]	ML	0.77	2	1		
					AE	0.72				
			SVD	[39]	ML	0.78	1			
					AE	0.74				
			Tensor factorization	[42]	ML	0.75	3			
					AE	0.72				
		User-based CF	–	[46]	ML	0.70	–	3		
					AE	0.68				
	Graph-based model	–	–	[50]	ML	0.68	–	–	2	
					AE	0.66				
	Rule-nased model	–	–	[56]	ML	0.66	–	–	4	
					AE	0.63				
Hybrid & combined approach	Ensemble model	–	–	[60]	ML	0.80	–	–	1	2
					AE	0.75				
	Ranking model	–	–	[63]	ML	0.78	–	–	2	
					AE	0.73				
Deep learning	Context-aware model	–	–	[68]	ML	0.77	–	–	8	1
					AE	0.73				
	Attention and memory network	–	–	[72]	ML	0.89	–	–	2	
					AE	0.89				
	Neural graph based	GNNs	–	[78]	ML	0.88	–	1	3	
					AE	0.86				
		Neural CF	–	[81]	ML	0.87	–	2		
					AE	0.85				
	Autoencoder	–	–	[83]	ML	0.83	–	–	5	
					AE	0.81				
	RNNs & LSTM	–	–	[22]	ML	0.79	–	–	7	
					AE	0.75				
	Sequence-aware	–	–	[91]	ML	0.84	–	–	4	
					AE	0.83				
	Convolutional operation	GCNs	–	[5]	ML	0.81	–	1	6	
					AE	0.79				
		CNNs	–	[21]	ML	0.80	–	2		
					AE	0.77				
	Self-attention mechanism	–	–	[103]	ML	0.91	–	–	1	
					AE	0.90				
Algorithmic and mathematical modeling	Probabilistic/ statistical model	–	–	[24]	ML	0.64	–	–	1	4
					AE	0.60				
	Clustering-based model	–	–	[110]	ML	0.62	–	–	2	
					AE	0.56				

method, and (4) the ranking of the methods.

• Figures 2 and 3 illustrate the individual MAP@10, MAP@30, and MAP@50 for each algorithm using MovieLense dataset (Fig. 2) and Amazon Electronics Dataset (Fig. 3). The algorithms in each figure are grouped based on the common methods they employ.

6.5 Discussion of experimental results

6.5.1 User and item similarity based methods

6.5.1.1 Content-based filtering[25]

The algorithm effectively suggests items across various categories but tends to recommend those similar to previously interacted items, suggesting a "filter bubble" with reduced exposure to diverse content. As the dataset size enlarges, computational time linearly rises, showing good scalability. However, its performance dips slightly with very large datasets, indicating a need for optimization in handling large big data. Compared to collaborative filtering models, this content-based approach excells in scenarios with sparse user data but is less adept at capturing complex user preferences. It effectively recommends items based on demographic data but struggles with items lacking historical data, highlighting the cold start problem. An
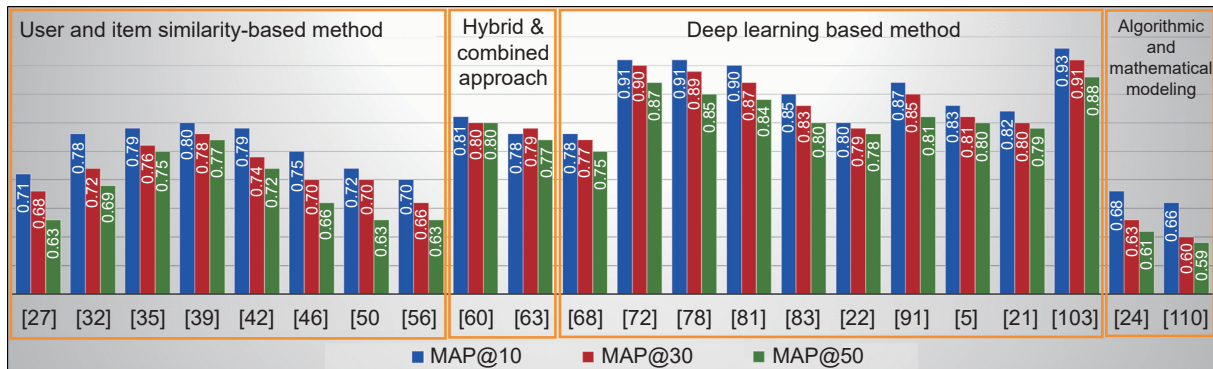
Fig. 2 Individual MAP@10, MAP@30, and MAP@50 for each algorithm using MovieLense dataset. The algorithms are grouped based on the common methods they employ.
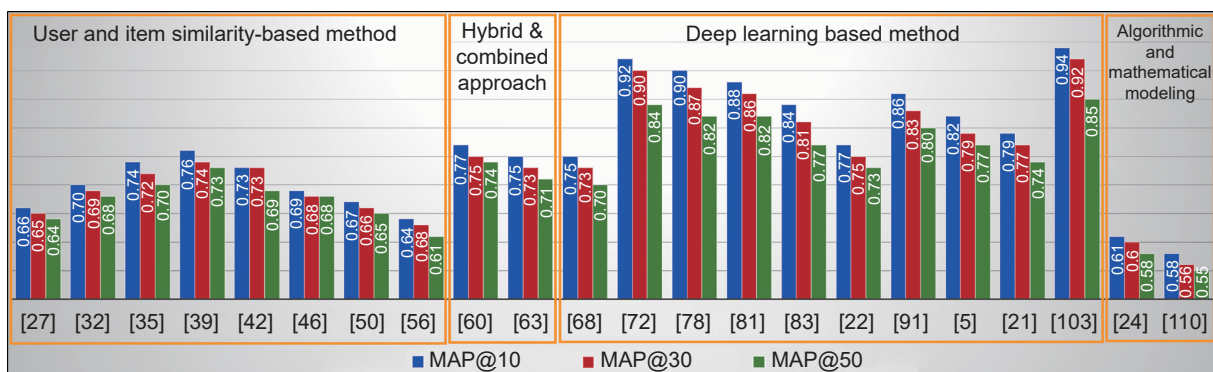

Fig. 3 Individual MAP@10, MAP@30, and MAP@50 for each algorithm using Amazon Electronics dataset. The algorithms are grouped based on the common methods they employ.

analysis shows a bias towards popular items, risking the underrepresentation of niche content. Longitudinal studies suggest users' content consumption becomes more homogenized over time.

6.5.1.2 Item-based CF[32]

Our investigation reveals a consistent decrement in performance concurrent with an augmentation of latent factors, reaching a plateau at 20 factors. This plateau suggests an optimal equilibrium between the complexity of the model and its efficacy. Comparatively, the algorithm surpasses most existing collaborative filtering algorithms in terms of precision. However, it does not match the performance of deep learning based algorithms, particularly in addressing cold-start situations. The quantity of latent factors emerges as a pivotal element influencing model efficiency. We observe a diminishing return on performance enhancement beyond the threshold of 50 factors, indicating an upper boundary for their practical application. Scalability assessments demonstrate that the algorithm upholds reasonable computation durations up to the threshold of 500 000 ratings.

Beyond this juncture, there is a marked escalation in computational demands. The algorithm exhibits commendable capabilities in generating recommendations for users with sparse interaction data, addressing a prevalent obstacle in recommendation systems. The algorithm's effectiveness diminishes for items with few ratings, highlighting the need for improved strategies to enhance recommendations for newly introduced items.

6.5.1.3 Matrix factorization[35, 39, 42]

In comparison to the other user and item similarity based algorithms, the algorithms based on matrix factorization demonstrate a markedly superior performance, particularly in scenarios characterized by high data sparsity. This superior performance implies that matrix factorization algorithms are more adept at identifying latent factors within user-item interactions. Our systematic experimentation indicates a pronounced sensitivity of these algorithms to hyperparameters, such as the number of latent factors and the regularization coefficient. Optimum results are achieved with approximately 100 latent factors, beyond which there is

a noticeable trend of diminishing returns and emerging risks of overfitting. The ability of these matrix factorization algorithms to adeptly manage sparse datasets, a prevalent issue in recommendation systems, is particularly notable, as evidenced by their enhanced performance in datasets exhibiting over about 60% sparsity. Scalability assessments reveal that although these algorithms effectively process larger datasets, this is accompanied by an increased computational duration, suggesting an intrinsic trade-off between accuracy and efficiency that must be weighed in practical applications. The overemphasis on regularization in the algorithm leads to underfitting, emphasizing the importance of careful parameter tuning. Addressing the cold start problem remains a major challenge for these algorithms.

6.5.1.4   User-based CF[46]

The algorithm demonstrates robust performance with regard to precision, underscoring its efficacy in accurately identifying items pertinent to users' interests. Although the recall rate is somewhat less impressive compared to precision, it remains significant, suggesting that there is potential for further enhancement in encompassing a broader spectrum of relevant items. The diversity score could be improved to ensure that users are exposed to a wider variety of items, potentially enhancing user satisfaction.

6.5.1.5   Graph-based models[50]

The graph-based algorithm exhibits noteworthy precision, surpassing several other algorithms in contexts characterized by sparse user-item interactions. This significant enhancement underscores the algorithm's proficiency in deciphering intricate relationships embedded within the dataset. In a comparative analysis with matrix factorization algorithms, the graph-based algorithm not only matches in precision but also excells in managing cold-start.

This superiority implies its enhanced capability in navigating scenarios involving new users or items with minimal historical data. However, scalability assessments indicate that while the algorithm demonstrates exceptional performance on smaller datasets, a discernible escalation in computational complexity arises as the dataset size expands, hinting at a possible limitation in its application to vast-scale recommendation systems without additional optimizations. Also, a slight dip in performance is observed in exceedingly dense datasets, signaling a necessity for more precise adjustments in such environments.

6.5.1.6   Rule-based models[56]

In comparison to the other algorithms that rely on user and item similarity, the rule-based algorithm demonstrates a lower efficacy in tailoring recommendations to individual preferences but offers faster response time. In scenarios where user preferences are explicitly defined, this algorithm surpasses collaborative filtering algorithms in achieving greater precision. It exhibits notable efficiency in contexts with a limited number of items, yet faces challenges in scaling proportionately with expanding inventories.

The algorithm maintains optimal resource usage up to a certain point; however, the complexity inherent in managing and updating an extensive rule set beyond this threshold adversely affects its performance. A significant challenge encountered is the upkeep of an extensive rule set, necessitating continual revisions to align with evolving user behaviors. The algorithm also exhibits inflexibility in adapting to emergent trends.

6.5.2   Deep learning based methods

6.5.2.1   Context-aware models[68]

The context-aware algorithm demonstrates potential in augmenting user engagement, notwithstanding its position as the least accurate among competing deep learning based algorithms. It exhibits a commendable balance between precision and recall metrics. Of particular note is the algorithm's about 20% improvement in precision over the algorithms based on user and item similarity, a testament to its effective utilization of contextual data. Upon evaluating the algorithm across diverse user scenarios, it is observed to perform exceptionally with users who have extensive interaction histories. Conversely, its efficacy diminishes for users with limited historical data, underscoring the necessity for an enhanced approach to address the "cold-start" problem.

6.5.2.2   Attention and memory network models[72]

The attention and memory networks algorithm manifests a marked enhancement in precision, outperforming all other deep learning methodologies except for the self-attention mechanisms algorithm. This advancement is highlighted by a 15% increment in precision rate, underscoring a heightened relevance in item selection for users. The algorithm's attention

component is instrumental in augmenting the relevance of recommendations. By assigning weights to user-item interactions according to their significance, the model is adept at concentrating on impactful factors, thereby yielding more tailored recommendations. This enhancement is largely attributed to the effective incorporation of attention mechanisms, which adeptly captures user preferences.

When compared with collaborative filtering and matrix factorization techniques, this algorithm exhibits superior prowess in both recall and precision metrics. This superiority implies that the integration of memory networks offers an improved approach to managing sparse data. It shows proficiency in addressing the cold start problem. Also, in terms of scalability, the algorithm demonstrates consistent performance, even with escalating dataset sizes, signifying its applicability in real-world scenarios. However, the algorithm's computational complexity presents a hurdle that necessitates further refinement. The algorithm exhibits limitations when dealing with extremely sparse datasets and shows a propensity for bias towards frequently encountered items.

6.5.2.3 Neural graph-based models[78, 81]

The neural graph-based algorithms have a significant advancement in the realm of recommendation systems, surpassing the capabilities of the matrix factorization algorithms. A notable enhancement of over 20% in precision is observed when these algorithms are applied to the MovieLens dataset. This improvement underscores the algorithms' proficiency in harnessing the graph structure to intricately map user-item interactions. A critical aspect of this success is attributed to the neural network layers, which adeptly capture the non-linear and multifaceted nature of these interactions.

Our analysis reveals that the performance of these algorithms is considerably influenced by the architectural choices, particularly the number of hidden layers and the neurons within each layer. The optimum architecture is identified as comprising three hidden layers. Expanding the architecture beyond this configuration does not result in substantial improvements. This phenomenon indicates a threshold beyond which further complexity in the model ceases to offer proportional benefits. A challenge encountered is the cold start problem. In terms of scalability, the algorithms demonstrate a consistent performance level

as the dataset sizes expand. When handling datasets exceeding one million user-item interactions, a marked escalation in computational demands is observed. The algorithms' performance also displays sensitivity to specific hyperparameters, notably the embedding size and the number of graph convolution layers. An optimal balance between precision and computational efficiency is achieved with an embedding size of 128 and three graph convolution layers.

6.5.2.4 Autoencoders-based models[8]

The autoencoder algorithm has exhibited promise in recommendation systems, particularly in managing datasets that are both sparse and intricate in nature. When applied to the MovieLens dataset, the algorithm attains an MAP of 0.83, signifying a commendably high level of precision in forecasting user ratings. In contrast, its application to the Amazon Reviews dataset yields an average MAP of 0.81, denoting a marginally reduced accuracy, which may be attributed to the dataset's larger size.

Compared to collaborative filtering algorithms, the autoencoder demonstrates an enhancement of over 15% in precision on the MovieLens dataset. The sparsity characteristic of the Amazon Reviews dataset presents a formidable challenge, resulting in a minor decline in metric performance. Nonetheless, the autoencoder's proficiency in learning complex, nonlinear representations is instrumental in discerning deeper user-item interactions, even in contexts of sparse data. The training of the autoencoder is executed with a learning rate of 0.001 and a batch size of 128. Increasing the batch size speeds up training but slightly reduces precision. Implementing dropout regularization prevents overfitting, especially for the denser MovieLens dataset.

6.5.2.5 Recurrent neural networks and LSTM models[22]

The RNN algorithm exhibits a superior enhancement in prediction accuracy compared to collaborative filtering algorithms. In the MovieLense dataset, the RNN achieves an over 10% improvement in precision. This indicates the RNN's proficiency in deciphering sequential user behavior patterns, leading to more accurate recommendations. When juxtaposed with CNN-based algorithms, RNN showcases comparable efficacy in sequential recommendation tasks, attributes to its adeptness in temporal sequence modeling, a feature not intrinsic to CNNs. However, this comes

with a trade-off, as RNN entails a notably higher computational complexity, with training durations approximately tripling those of matrix factorization algorithms. This balance between precision and efficiency is pivotal for real-time recommendation systems.

Upon segment-based analysis, the RNN is especially potent for users with varied and substantial interaction histories. In contrast to other deep learning algorithms, the RNN, while slightly less precise, offers greater interpretability, shedding light on the evolution of user preferences over time. Optimizing the RNN's performance necessitates extensive hyperparameter adjustments. The number of hidden layers and the choice of recurrent unit profoundly influence the results, with optimal configurations varying based on dataset characteristics like sparsity and temporal dynamics. Another limitation of the RNN is its susceptibility to the sequence of user interactions, which might not always align with true user inclinations. The scalability of the model remains a challenge, particularly in environments encompassing millions of users and items, constrained by computational limitations.

6.5.2.6 Convolutional operations-based models[5, 21]

The algorithms leveraging convolutional operations exhibit noteworthy enhancements in precision. These improvements are principally attributed to the algorithms' proficient ability to discern and utilize spatial hierarchies within the data, thereby yielding more precise and contextually relevant recommendations. Our empirical studies further highlight the algorithms' adeptness in managing large datasets. Notably, as the data volume escalates, the algorithms demonstrate admirable scalability, sustaining their efficiency without a corresponding increase in computational burden.

However, these algorithms necessitate an approximate 30% increase in computational time compared to most collaborative filtering algorithms, owing to the integrated convolutional operations. Nevertheless, this increase in resource consumption is counterbalanced by an enhancement in the quality of recommendations. Rigorous testing across datasets of diverse sizes and characteristics confirms the algorithms' consistent efficacy, underscoring their robustness and scalability. This aspect is particularly vital in practical scenarios, where the volume and diversity of data can vary significantly.

6.5.2.7 Self-attention-based model[103]

The empirical evidence from our study clearly indicates that the algorithm employing the self-attention mechanism has outperformed all other examined algorithms in terms of precision. This enhanced performance can be primarily attributed to the mechanism's sophisticated capability to discern and interpret intricate user-item interactions and sequential trends within the dataset. The algorithm is notably proficient in managing sparse datasets, a frequent challenge in recommendation systems, by effectively addressing long-range dependencies. In terms of scalability, the self-attention algorithm exhibits linear progression in performance with the escalation in dataset size.

Remarkably, its accuracy remains consistent and commendable, even when confronted with a tenfold augmentation in data volume. This feature is particularly vital in real-world applications, where enormous data volumes are the norm. The algorithm's prowess in analyzing and anticipating user behavioral patterns is significantly bolstered by the self-attention mechanism. Despite its complex structure, the algorithm maintains computational efficiency, achieving reduced latency in generating recommendations when compared to the RNN-based algorithm. However, the algorithm presents certain limitations, particularly in scenarios involving highly diverse user interests.

6.5.3 Algorithmic and mathematical modelling[24, 110]

The analysis of the clusters generated by these algorithms reveals uniformity within the groups. This uniformity strongly indicates the algorithms' efficacy in accurately discerning user preferences. Impressively, the algorithms maintain robust functionality, even when applied to larger datasets, consistently demonstrating quick response time coupled with minimal computational demands. A key feature of these algorithms is their capacity to create coherent and meaningful clusters, as evidenced by the high intra-cluster similarity. Notably, the algorithms exhibit enhanced proficiency in processing sparse data and accommodating diverse user profiles, thereby yielding more tailored and pertinent recommendations. The algorithm displays remarkable scalability, sustaining stable performance despite increases in dataset size.

7 Challenges and Limitations of Current Methods for Big Data in Recommendation Systems

Content-based filtering methods: (a) Challenges: Difficulty in handling unstructured data, reliance on item descriptions, and potential for limited diversity in recommendations; and (b) Limitations: Tends to recommend items similar to those the user has already consumed, leading to a lack of novelty.

Item-based collaborative filtering methods: (a) Challenges: Scalability issues with large datasets and sparsity of user-item interactions; and (b) Limitations: May not perform well when new items are added, as they lack user interactions.

PMF methods: (a) Challenges: Computational complexity and difficulty in incorporating context or content information; and (b) Limitations: Can struggle with very sparse data and cold start problem for new users or items.

SVD methods: (a) Challenges: Computationally intensive, especially for large matrices; and (b) Limitations: Less effective when dealing with highly sparse matrices and not suitable for dynamic datasets where user preferences change frequently.

Tensor factorization methods: (a) Challenges: High computational cost and complexity in model tuning; and (b) Limitations: May overfit on smaller datasets and struggle with scalability.

User-based collaborative filtering methods: (a) Challenges: Scalability with a large number of users and sparsity in user-item interactions; and (b) Limitations: Often less accurate than item-based methods and sensitive to user profile changes.

Graph-based models: (a) Challenges: Computational complexity and difficulty in handling dynamic changes in the data; and (b) Limitations: Can be less effective for sparse datasets and may require extensive parameter tuning.

Rule-based models: (a) Challenges: Rigid structure that may not adapt well to changes in user behavior; and (b) Limitations: May not capture complex patterns or relationships in the data.

Ensemble models: (a) Challenges: Complexity in combining various models effectively and in tuning hyperparameters; and (b) Limitations: Risk of overfitting and high computational cost.

Ranking models: (a) Challenges: Difficulty in accurately modeling user preferences and handling large item sets; and (b) Limitations: May not account for long-term user preferences.

Context-aware models: (a) Challenges: Incorporating contextual information effectively and dealing with dynamic contexts; and (b) Limitations: Complexity in model design and potential privacy concerns.

GNNs methods: (a) Challenges: Scalability issues and computational complexity; and (b) Limitations: Can struggle with very sparse graphs and require significant training data.

NCF methods: (a) Challenges: Require large amounts of data and computational resources; and (b) Limitations: Can overfit smaller datasets and may not generalize well across different domains.

Autoencoders methods: (a) Challenges: Sensitive to hyperparameter settings and require significant training data; and (b) Limitations: Can struggle with capturing complex user-item interactions.

RNNs and LSTM methods: (a) Challenges: Difficulty in capturing long-term dependencies and computational intensity; and (b) Limitations: Can be prone to overfitting and struggle with very long sequences.

Sequence-aware models: (a) Challenges: Complexity in capturing the temporal dynamics of user interactions; and (b) Limitations: Can be computationally expensive and require extensive tuning.

GCNs methods: (a) Challenges: High computational cost and complexity in handling large graphs; and (b) Limitations: May require extensive domain knowledge for effective implementation.

CNNs methods: (a) Challenges: Requirement for large training datasets and computational resources; and (b) Limitations: May not be suitable for data that lacks spatial or temporal structure.

DRL methods: (a) Challenges: Require extensive training data and computational resources; and (b) Limitations: Can be difficult to converge and sensitive to hyperparameter settings.

Self-attention mechanisms methods: (a) Challenges: High computational complexity and memory requirements; and (b) Limitations: Can struggle with very long sequences and require careful tuning.

Probabilistic and statistical models methods: (a) Challenges: Complexity in model formulation; and (b)

Limitations: Can be less effective with non-linear relationships in data.

8 Real-World Applications and Case Studies of the Methods for Big Data in Recommendation Systems

We provide in this section examples that illustrate the diverse and innovative ways, in which big data and various recommendation system methods are being applied across industries to enhance user experience and business efficiency.

Netflix's content-based movie suggestions[113]: Netflix uses content-based filtering to recommend movies and TV shows by analyzing the properties of the content (like genre, actors, director) and matching them with a user's past preferences.

Amazon's item-based collaborative filtering recommendations[114]: Amazon employs item-based collaborative filtering to suggest products based on similarities between items and user interactions, such as purchasing history and item ratings.

Spotify's PMF music recommendations[115]: Spotify leverages probabilistic matrix factorization to handle scalability and sparsity in its large datasets, providing personalized music and playlist recommendations.

SVD for E-commerce product recommendation[116]: Many E-commerce platforms use SVD-based recommendation systems to suggest products to users. This includes not only large players like Amazon but also smaller E-commerce sites. These systems analyze user behavior data, including views, clicks, and purchases, to recommend products that users are likely to buy.

Tensor factorization in E-commerce personalization[117]: Online retail platforms use tensor factorization to analyze multidimensional data (user, item, time) for personalized shopping experiences.

LinkedIn's user-based collaborative filtering recommendations[118]: LinkedIn uses user-based collaborative filtering to suggest new connections by looking at the similarities between users and their networks

Google's knowledge graph for search recommendations[119]: Google uses graph-based models to enhance its search engine, offering relevant search suggestions and semantic search results.

Rule-based recommendations in travel (Booking.com, Airbnb)[120]: In these sectors, recommendation systems might use rules based on travel history, search behavior, and preferences. For example, if a user often books seaside hotels, the system might prioritize similar destinations or properties near the beach in future recommendations.

Google Play's ensemble learning recommendations[121]: Google Play uses ensemble learning in its recommendation systems to suggest apps, games, books, and movies. The system combines user data with app attributes and contextual information, using a variety of machine learning models to create personalized recommendations.

YouTube's recommendation-driven video ranking[122]: YouTube utilizes ranking models to order video recommendations based on predicted user engagement and relevance.

Uber's context-aware restaurant recommendations in Uber Eats[123]: Uber Eats uses context-aware models to recommend restaurants based on the user's location, time, and previous orders.

Friend recommendations via GNNs in social media[124]: Social networks like Facebook could use GNNs to suggest friends by analyzing the complex connections in the social graph.

NCF in E-commerce[125]: NCF methods can be applied in E-commerce platforms for personalized product recommendations based on deep learning user-item interactions.

Autoencoder-powered content discovery (e.g., Pinterest)[126]: Platforms like Pinterest could use autoencoders for unsupervised learning of user preferences and content features to recommend similar items.

RNN and LSTM for sequential recommendation systems[127]: Platforms like Netflix may use RNN and LSTM for sequence-aware recommendations, such as suggesting the next episode in a TV.

Sequence-aware music recommendations (e.g., Spotify)[128]: Services like Spotify use sequence-aware models to recommend songs that fit the context of the user's current playlist.

GCNs in PinSage's producer-consumer recommendations [129]: PinSage is built on the concept of GNN, which allows it to efficiently process information structured in the form of a graph. This is particularly suitable for platforms like Pinterest, where items (like pins) can be naturally represented in a graph

structure, with nodes representing the items and edges representing relationships between them.

Visual recommendations with CNNs in fashion retail[130]**:** Fashion retailers like ASOS may use CNNs for visual search and recommendations, analyzing images to suggest visually similar items.

E-commerce strategies with DRL[131]**:** E-commerce sites might use deep reinforcement learning for inventory recommendations based on dynamic market conditions.

Product review analysis with self-attention mechanisms[132]**:** Recommendation systems could use models with self-attention mechanisms, like Transformer models, to analyze and summarize product reviews for better recommendations.

Marketing customer segmentation via clustering methods[133]**:** Clustering-based models are used in marketing for segmenting customers into groups for targeted product recommendations.

9 Future Directions for Big Data in Recommendation Systems: Improvement, Ethics, and Bias Perspectives

9.1 Improvement perspectives and directions

9.1.1 User and item similarity based methods

• **Enhanced personalization through deep learning:** Big data can enable more sophisticated deep learning models that understand nuanced user preferences and item characteristics. These models can learn complex patterns and relationships, improving the accuracy of recommendations based on user and item similarities.

• **Dynamic user profiling:** As big data continuously evolves, recommendation systems can dynamically update user profiles. This means that recommendations can adapt in real-time to changes in user behavior, preferences, or circumstances, maintaining relevance and engagement.

• **Cross-domain recommendations:** Leveraging big data from various domains can enable cross-domain recommendation systems. For example, understanding a user's preferences in music might inform recommendations in movies or books, assuming certain overlap in taste.

• **Improved item similarity detection:** Big data can enhance item similarity algorithms by incorporating a wider range of attributes and metadata. This can lead to more nuanced and accurate item categorization, which in turn leads to more relevant recommendations.

• **Scalability and performance optimization:** With the growing size of datasets, optimizing the scalability and performance of similarity-based methods becomes crucial. Future developments may focus on efficient algorithms that can handle large-scale data without compromising accuracy.

• **Integration of contextual data:** Beyond just user and item data, incorporating contextual information (like location, time, social context) can significantly refine recommendations. Big data enables the aggregation and analysis of such contextual details.

• **Visual and sensory data utilization:** Future recommendation systems might incorporate more visual and sensory data (like images, videos, and sounds) to enhance item similarity measures, especially in fields like fashion, art, and entertainment.

• **Interactive and adaptive systems:** The integration of AI and big data can lead to more interactive systems where users can provide real-time feedback, further refining the recommendation process based on user and item similarity.

• **Blockchain for transparency and security:** Utilizing blockchain technology can increase transparency and security in how data is used and shared, particularly in CF methods involving user and item similarities.

9.1.2 Hybrid and combined approaches

• **Enhanced personalization through hybrid models**: By combining content-based and collaborative filtering methods, hybrid approaches can offer more personalized recommendations. Future systems may utilize advanced machine learning algorithms to understand user preferences more deeply, even in sparse data scenarios.

• **Integration of contextual data:** Hybrid systems can be enriched by integrating contextual information like time, location, or specific user circumstances. This context awareness can lead to more relevant and timely recommendations, improving user experience.

• **Utilization of deep learning techniques:** The integration of deep learning techniques in hybrid recommendation systems can lead to more sophisticated feature extraction and better handling of unstructured data, like images and text. This could result in accurate predictions.

• **Cross-domain recommendations:** Future hybrid systems may effectively leverage cross-domain data, offering users recommendations that span different

types of services or products. This approach can lead to discovering new user interests and broadening the scope of recommendations.

• **Real-time data processing:** With the increasing velocity of big data, real-time recommendation systems will become more prominent. Hybrid systems could combine real-time analytics with historical data analysis to provide immediate and relevant recommendations.

• **Interactive and adaptive systems:** Future recommendation systems might become more interactive, allowing users to provide feedback that instantly adjusts the recommendation logic. This adaptability can lead to a continuously improving and personalizing recommendation experience.

• **Explainable AI in recommendations:** There is a growing demand for explainable AI, where systems can provide reasons behind their recommendations. Future hybrid models could incorporate explainability aspects, making recommendations more transparent and trustworthy.

• **Integration with emerging technologies:** Hybrid recommendation systems could be integrated with emerging technologies like Augmented Reality (AR), Virtual Reality (VR), and the Internet of Things (IoT), offering unique and immersive recommendation experiences.

9.1.3 Deep learning methods

• **Enhanced personalization:** Deep learning methods can analyze vast datasets to understand nuanced user preferences, leading to highly personalized recommendations. This goes beyond traditional recommendation systems by considering a wider range of user behaviors and interactions.

• **Context-aware recommendations**: Future recommendation systems may leverage deep learning to better understand the context in which recommendations are made. This includes recognizing the user's current environment, time, mood, or even social setting, leading to more appropriate and timely suggestions.

• **Sophisticated content analysis:** Deep learning excels in interpreting complex content like images, videos, and text. This capability can be utilized to develop recommendation systems that understand content at a deeper level, providing more relevant suggestions based on the content's inherent qualities, rather than just metadata.

• **Predictive user modeling:** By harnessing the predictive power of deep learning with big data, recommendation systems can anticipate user needs and preferences, potentially recommending items even before the user explicitly expresses a desire for them.

• **Overcoming data sparsity and cold start challenges:** Deep learning can effectively deal with sparse data scenarios, such as when new users or items have limited interaction history. Advanced models can infer preferences from limited data, significantly improving the user experience from the start.

• **Continuous learning and evolution:** Deep learning models can be designed to continuously learn and adapt over time, ensuring that recommendations remain relevant as user preferences and behaviors evolve.

• **Multi-modal data integration:** Future recommendation systems might integrate and analyze data from multiple sources and types — including text, images, audio, and sensor data — to create a more comprehensive understanding of user preferences.

• **Resistance to manipulation:** Deep learning can help in detecting and countering fraudulent activities like fake reviews and ratings, ensuring the integrity and trustworthiness of the recommendations.

• **Enhanced collaborative filtering:** Incorporating deep learning in collaborative filtering methods can lead to a more nuanced understanding of user-item interactions, improving the accuracy of recommendations.

• **Real-time recommendation capabilities:** The processing power of deep learning models could enable real-time analysis of user data, allowing for instant, dynamic recommendations based on current user activities.

• **Integration with emerging technologies:** Deep learning methods could be synergized with emerging technologies like augmented reality and virtual reality for creating immersive and interactive recommendation experiences.

9.1.4 Algorithmic and mathematical modelling methods

• **Advanced statistical models:** Future recommendation systems could employ more sophisticated statistical models to analyze big data. Techniques like Bayesian networks and Markov decision processes might be used to model user preferences and predict future behavior more

accurately.

- **Complex network analysis:** Utilizing the principles of network theory, recommendation systems can analyze the intricate web of relationships between users and items. By applying measures like centrality, clustering coefficients, and community detection, these systems can provide more nuanced and relevant recommendations.

- **Optimization techniques:** The use of mathematical optimization, such as linear programming, integer programming, and constraint programming, can enhance the efficacy of recommendation systems. These methods can help in balancing multiple objectives, like maximizing relevance while ensuring diversity in the recommendations.

- **Graph-based mathematical models:** Developing new graph-based algorithms, perhaps inspired by recent advances in graph theory, could lead to more effective ways of understanding and leveraging the connections between users and items in recommendation systems.

- **Quantitative behavioral models:** Mathematical models that quantify and predict human behavior, such as those based on theories of consumer choice or decision-making processes, can be integrated into recommendation systems. These models can provide deeper insights into why users prefer certain items, leading to more effective recommendations.

- **Game theory and mechanism design:** Employing concepts from game theory and mechanism design could enhance the way recommendation systems handle interactions and competitions among multiple stakeholders (e.g., users, providers, advertisers).

- **Differential equations for dynamic modeling:** Using differential equations to model the dynamic and evolving nature of user preferences and item popularity can offer a realistic representation of the recommendation.

- **Multi-Criteria Decision Analysis (MCDA):** MCDA methods can be applied to consider various criteria that users might have for recommendations, providing a more holistic approach to generating suggestions.

- **Quantum algorithms for data processing**: The potential of quantum computing in processing large datasets could revolutionize recommendation systems. Quantum algorithms could handle complex computations much faster than traditional methods, allowing for real-time analysis of vast amounts of data.

- **Integration of explainable AI (XAI):** With the increasing complexity of algorithms, incorporating XAI into recommendation systems will be crucial for transparency and user trust. Users could understand the rationale behind recommendations, which is particularly important in sensitive areas like healthcare or finance.

9.2 Ethics and bias perspectives

9.2.1 Potential biases in recommendation systems

- **Data bias**: Biases in data collection can lead to skewed recommendations, often reflecting societal biases or underrepresentation of certain groups. To address this, future systems might focus on proactive measures such as inclusive data sourcing and enhanced demographic analysis to ensure a wide representation of diverse populations. This could involve partnerships with organizations representing marginalized communities to gather more inclusive datasets.

- **Algorithmic bias:** Algorithms designed for specific goals, like maximizing engagement, can inadvertently promote polarizing content or misinformation. Future strategies could include the implementation of ethical guidelines in algorithm design, emphasizing the balance between user engagement and the promotion of diverse, accurate content. Additionally, interdisciplinary teams comprising ethicists, sociologists, and data scientists could be formed to oversee the ethical implications of algorithmic decisions.

9.2.2 Mitigating strategies and best practices

- **Diverse data sets:** Using diverse and representative datasets helps reduce bias. This could be further enhanced by employing techniques like synthetic data generation to fill gaps in underrepresented areas, ensuring a broader spectrum of user preferences and behaviors is captured.

- **Regular algorithm auditing:** Audits can identify and mitigate biases by testing algorithms across various groups and scenarios. Future systems could adopt continuous monitoring frameworks, using AI-driven tools to detect and address bias in real-time, rather than relying on periodic audits. This proactive approach can quickly adapt to emerging biases or changes in societal norms.

- **Fairness algorithms:** Developing algorithms to promote fairness can counteract biases in data and

predictions. Future perspectives could see the integration of fairness as a core component in the algorithmic design process, rather than an afterthought. This might involve the use of fairness metrics that are regularly updated to reflect evolving societal standards and values.

• **User feedback mechanisms:** Allowing user feedback and control over data usage can lead to more balanced, user-centric systems. In the future, systems might offer more granular control to users over their data, including the ability to opt-in or out of specific data collection practices. Additionally, incorporating direct user feedback into the recommendation algorithm itself can provide a more dynamic and responsive system that aligns closely with individual preferences and values.

10   Conclusion

This survey paper has addressed the existing gaps in the literature on big data algorithms in recommendation systems by providing a detailed and comprehensive analysis coupled with an innovative hierarchical taxonomy. The taxonomy, characterized by its tri-level hierarchy, categorizes algorithms into four main analysis types—user and item similarity based methods, hybrid and combined approaches, deep learning and algorithmic methods, and mathematical modeling methods—further broken down into sub-categories and techniques. This structured framework facilitates a clearer understanding of the relationships and distinctions among various algorithms.

The survey's unique contribution lies in its methodical approach, combining empirical and experimental evaluations to rank these algorithms within their respective categories and techniques. This dual evaluation system not only distinguishes the techniques based on specific criteria but also enables a nuanced comparison within the same category and technique, offering valuable insights into their effectiveness.

By addressing the previous shortcomings in the literature—namely, the lack of an up-to-date, detailed overview and the broad, imprecise categorization of algorithms—this paper paves the way for more accurate and systematic research in the field. The findings from this survey are poised to aid future research, providing a clearer direction for the development and application of big data techniques in

recommendation systems, while highlighting areas ripe for future exploration and advancement.

Here is an outline of the main findings from our experimental evaluation:

• Content-based filtering shows good scalability and effectiveness in sparse data scenarios but tends to create a "filter bubble".

• Item-based collaborative filtering excells in precision but struggles with cold-start problems and data scalability.

• Matrix factorization algorithms outperform others in high data sparsity situations, though they are sensitive to hyperparameter settings and faced scalability challenges.

• User-based collaborative filtering and graph-based models demonstrate high precision, especially in sparse interaction contexts, but have limitations in scalability and dataset density.

• Rule-based models are fast and precise in specific scenarios but less effective with large and diverse datasets.

• Context-aware models, attention and memory network models, and neural graph-based models show significant improvements in precision and handling sparse datasets.

• Autoencoders and RNNs are effective in learning complex patterns and sequential user behaviors, though they face challenges in computational complexity and scalability.

• Convolutional operations-based models balance precision with scalability but require higher computational resources.

• Self-attention models excell in precision and scalability but have limitations in handling diverse user interests.

• Algorithmic and mathematical modeling algorithms efficiently process large, sparse datasets, maintaining quick response times and minimal computational demands. Each algorithmic approach exhibits unique advantages in specific contexts, underscoring the need for careful selection and optimization based on the specific requirements of the recommendation system being developed.

References

[1]   J. Nie, Z. Zhao, L. Huang, W. Nie, and Z. Wei, Cross-domain recommendation via user-clustering and multidimensional information fusion, *IEEE Trans. Multimedia*, vol. 25, pp. 868–880, 2023.

[2] G. Zhao, X. Lei, X. Qian, and T. Mei, Exploring users' internal influence from reviews for social recommendation, *IEEE Trans. Multimedia*, vol. 21, no. 3, pp. 771–781, 2019.

[3] J. Xin, Z. Wang, L. Qu, and G. Wang, Elastic extreme learning machine for big data classification, *Neurocomputing*, vol. 149, pp. 464–471, 2015.

[4] C. Yang, E. Kristiani, and Y. L. Chang, Big data and machine learning driven bioprocessing – Recent trends and critical analysis, *Bioresour. Technol.*, vol. 372, p. 128625, 2023.

[5] J. Yu, H. Yin, J. Li, M. Gao, Z. Huang, and L. Cui, Enhancing social recommendation with adversarial graph convolutional networks, *IEEE Trans. Knowledge Data Eng.*, vol. 34, no. 8, pp. 3727–3739, 2022.

[6] B. A. Hammou, A. A. Lahcen, and S. Mouline, A distributed group recommendation system based on extreme gradient boosting and big data technologies, *Appl. Intell.*, vol. 49, no. 12, pp. 4128–4149, 2019.

[7] R. A. Zayed, L. F. Ibrahim, H. A. Hefny, and H. A. Salman, Shilling attacks detection in collaborative recommender system: Challenges and promise, in *Proc. Workshops of the 34th Int. Conf. Advanced Information Networking and Applications*, Caserta, Italy, 2020, pp. 429–439.

[8] R. Ranjan, O. Rana, S. Nepal, M. Yousif, P. James, Z. Wen, S. Barr, P. Watson, P. P. Jayaraman, D. Georgakopoulos, et al., The next grand challenges: Integrating the Internet of Things and data science, *IEEE Cloud Comput.*, vol. 5, no. 3, pp. 12–26, 2018.

[9] K. Taha and R. Elmasri, XCDSearch: An XML context-driven search engine, *IEEE Trans. Knowledge Data Eng.*, vol. 22, no. 12, pp. 1781–1796, 2010.

[10] W. Dai, Q. Zhang, W. Pan, and Z. Ming, Transfer to rank for top-N recommendation, *IEEE Trans. Big Data*, vol. 6, no. 4, pp. 770–779, 2020.

[11] K. Taha, D. Homouz, H. Al Muhairi, and Z. Al Mahmoud, GRank: A middleware search engine for ranking genes by relevance to given genes, *BMC Bioinformatics*, vol. 14, p. 251, 2013.

[12] G. Yue, R. Xiao, Z. Zhao, and C. Li, AF-GCN: Attribute-fusing graph convolution network for recommendation, *IEEE Trans. Big Data*, vol. 9, no. 2, pp. 597–607, 2023.

[13] W. Huang, J. Liu, T. Li, T. Huang, S. Ji, and J. Wan, FedDSR: Daily schedule recommendation in a federated deep reinforcement learning framework, *IEEE Trans. Knowledge Data Eng.*, vol. 35, no. 4, pp. 3912–3924, 2023.

[14] J. M. Shapiro, Smart cities: Quality of life, productivity, and the growth effects of human capital, *Rev. Econ. Statist.*, vol. 88, no. 2, pp. 324–335, 2006.

[15] Y. Cheng, H. Chen, L. Xu, K. Chen, X. Wang, and Z. Huang, Exploring the role of big data algorithm recommendation in smart cities-taking book recommendation as an example, in *Proc. 2023 Int. Conf.*

IT Innovation and Knowledge Discovery (ITIKD), Manama, Bahrain, 2023, pp. 1–5.

[16] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, Using collaborative filtering to weave an information tapestry, *Commun. ACM*, vol. 35, no. 12, pp. 61–70, 1992.

[17] G. Adomavicius and A. Tuzhilin, Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, *IEEE Trans. Knowledge Data Eng.*, vol. 17, no. 6, pp. 734–749, 2005.

[18] Y. Koren, R. Bell, and C. Volinsky, Matrix factorization techniques for recommender systems, *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[19] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver, Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering, in *Proc. 4th ACM Conf. Recommender Systems*, Barcelona, Spain, 2010, pp. 79–86.

[20] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[21] D. Liu, J. Li, B. Du, J. Chang, R. Gao, and Y. Wu, A hybrid neural network approach to combine textual information and rating information for item recommendation, *Knowledge Inf. Syst.*, vol. 63, no. 3, pp. 621–646, 2021.

[22] J. Wang, N. Wu, and W. X. Zhao, Personalized route recommendation with neural network enhanced search algorithm, *IEEE Trans. Knowledge Data Eng.*, vol. 34, no. 12, pp. 5910–5924, 2022.

[23] X. Sun, H. Cheng, J. Li, B. Liu, and J. Guan, All in one: Multi-task prompting for graph neural networks, in *Proc. 29th ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, Long Beach, CA, USA, 2023, pp. 2120–2131.

[24] N. Liu and J. Zhao, Recommendation system based on deep sentiment analysis and matrix factorization, *IEEE Access*, vol. 11, pp. 16994–17001, 2023.

[25] J. Sun, L. Gao, X. Shen, S. Liu, R. Liang, S. Du, and S. Liu, Separated graph neural networks for recommendation systems, *IEEE Trans. Ind. Inf.*, vol. 19, no. 1, pp. 382–393, 2023.

[26] X. Sun, H. Cheng, B. Liu, J. Li, H. Chen, G. Xu, and H. Yin, Self-supervised hypergraph representation learning for sociological analysis, *IEEE Trans. Knowledge Data Eng.*, vol. 35, no. 11, pp. 11860–11871, 2023.

[27] S. Hu, A. Kumar, F. Al-Turjman, S. Gupta, S. Seth, and Shubham, Reviewer credibility and sentiment analysis based user profile modelling for online product recommendation, *IEEE Access*, vol. 8, pp. 26172–26189, 2020.

[28] W. Wang, T. Tang, F. Xia, Z. Gong, Z. Chen, and H. Liu, Collaborative filtering with network representation learning for citation recommendation, *IEEE Trans. Big Data*, vol. 8, no. 5, pp. 1233–1246, 2022.

[29] S. Jiang, X. Qian, T. Mei, and Y. Fu, Personalized travel sequence recommendation on multi-source big social

media, *IEEE Trans. Big Data*, vol. 2, no. 1, pp. 43–56, 2016.

[30] D. Lian, Y. Ge, F. Zhang, N. J. Yuan, X. Xie, T. Zhou, and Y. Rui, Scalable content-aware collaborative filtering for location recommendation, *IEEE Trans. Knowledge Data Eng.*, vol. 30, no. 6, pp. 1122–1135, 2018.

[31] Y. Liu, C. Yang, J. Ma, W. Xu, and Z. Hua, A social recommendation system for academic collaboration in undergraduate research, *Expert Syst.*, vol. 36, no. 2, p. e12365, 2019.

[32] C. Yan, X. Cui, L. Qi, X. Xu, and X. Zhang, Privacy-aware data publishing and integration for collaborative service recommendation, *IEEE Access*, vol. 6, pp. 43021–43028, 2018.

[33] J. Sun and Y. Zhang, Recommendation system with biclustering, *Big Data Mining and Analytics*, vol. 5, no. 4, pp. 282–293, 2022.

[34] C. Zan, Development of e-commerce big data model based on machine learning and user recommendation algorithm, *Int. J. Syst. Assur. Eng. Manag.*, doi: 10.1007/s13198-023-02157-y.

[35] B. Yi, X. Shen, H. Liu, Z. Zhang, W. Zhang, S. Liu, and N. Xiong, Deep matrix factorization with implicit feedback embedding for recommendation system, *IEEE Trans. Ind. Inf.*, vol. 15, no. 8, pp. 4591–4601, 2019.

[36] M. Hu, D. Wu, R. Wu, Z. Shi, M. Chen, and Y. Zhou, *RAP*: A light-weight privacy-preserving framework for recommender systems, *IEEE Trans. Serv. Comput.*, vol. 15, no. 5, pp. 2969–2981, 2022.

[37] L. Yao, X. Wang, Q. Z. Sheng, B. Benatallah, and C. Huang, Mashup recommendation by regularizing matrix factorization with API co-invocations, *IEEE Trans. Serv. Comput.*, vol. 14, no. 2, pp. 502–515, 2021.

[38] X. Shen, B. Yi, H. Liu, W. Zhang, Z. Zhang, S. Liu, and N. Xiong, Deep variational matrix factorization with knowledge embedding for recommendation system, *IEEE Trans. Knowledge Data Eng.*, vol. 33, no. 5, pp. 1906–1918, 2021.

[39] J. Guan, X. Huang, and B. Chen, Community-aware social recommendation: A unified SCSVD framework, *IEEE Trans. Knowledge Data Eng.*, vol. 35, no. 3, pp. 2379–2393, 2023.

[40] D. Lian, X. Xie, and E. Chen, Discrete matrix factorization and extension for fast item recommendation, *IEEE Trans. Knowledge Data Eng.*, vol. 33, no. 5, pp. 1919–1933, 2021.

[41] J. Jiao, X. Zhang, F. Li, and Y. Wang, A novel learning rate function and its application on the SVD++ recommendation algorithm, *IEEE Access*, vol. 8, pp. 14112–14122, 2020.

[42] V. Ioannidis, A. Zamzam, G. Giannakis, and N. Sidiropoulos, Coupled graphs and tensor factorization for recommender systems and community detection, *IEEE Trans. Knowledge Data Eng.*, vol. 33, no. 3, pp. 909–920, 2021.

[43] S. Li, J. Wen, F. Luo, and G. Ranzi, Time-aware QoS prediction for cloud service recommendation based on matrix factorization, *IEEE Access*, vol. 6, pp. 77716–77724, 2018.

[44] S. Meng, H. Wang, Q. Li, Y. Luo, W. Dou, and S. Wan, Spatial-temporal aware intelligent service recommendation method based on distributed tensor factorization for big data applications, *IEEE Access*, vol. 6, pp. 59462–59474, 2018.

[45] Y. Yang, L. Han, Z. Gou, B. Duan, J. Zhu, and H. Yan, Tagrec-CMTF: Coupled matrix and tensor factorization for tag recommendation, *IEEE Access*, vol. 6, pp. 64142–64152, 2018.

[46] L. Qi, X. Xu, X. Zhang, W. Dou, C. Hu, Y. Zhou, and J. Yu, Structural balance theory-based e-commerce recommendation over big rating data, *IEEE Trans. Big Data*, vol. 4, no. 3, pp. 301–312, 2018.

[47] X. Wu and Z. Wu, Application of big data search based on collaborative filtering algorithm in cross-border e-commerce product recommendation, *Soft Comput.*, doi: 10.1007/s00500-023-08643-6.

[48] M. Hong and J. Jung, Towards social big data-based affective group recommendation, *Mobile Network. Appl.*, vol. 23, no. 4, pp. 1111–1122, 2018.

[49] H. Chen, Y. Li, X. Sun, G. Xu, and H. Yin, Temporal meta-path guided explainable recommendation, in *Proc. 14th ACM Int. Conf. Web Search and Data Mining*, Virtual Event, 2021, pp. 1056–1064.

[50] L. Qi, Q. He, F. Chen, X. Zhang, W. Dou, and Q. Ni, Data-driven web APIs recommendation for building web applications, *IEEE Trans. Big Data*, vol. 8, no. 3, pp. 685–698, 2022.

[51] X. Zhou, W. Liang, S. Huang, and M. Fu, Social recommendation with large-scale group decision-making for cyber-enabled online service, *IEEE Trans. Comput. Soc. Syst.*, vol. 6, no. 5, pp. 1073–1082, 2019.

[52] Z. Peng, Y. Yang, L. Yang, and K. Chen, Federated meta embedding concept stock recommendation, *IEEE Trans. Big Data*, doi: 10.1109/TBDATA.2022.3214622.

[53] W. Zhu, Research on multi-source mobile commerce service recommendation model of data fusion based on tree network, *Concurrency Computat. Pract. Exper.*, vol. 34, no. 13, p. e5862, 2022.

[54] R. Song, T. Li, X. Dong, and Z. Ding, An effective points of interest recommendation approach based on embedded meta-path of spatiotemporal data, *Expert Syst.*, vol. 40, no. 2, p. e13145, 2023.

[55] B. Wang, S. Song, S. Liu, and X. Deng, MultiFDF: Multi-community clustering for fairness-aware recommendation, *IEEE Trans. Comput. Soc. Syst.*, vol. 10, no. 6, pp. 2959–2970, 2023.

[56] I. Varlamis, C. Sardianos, C. Chronis, G. Dimitrakopoulos, Y. Himeur, A. Alsalemi, F. Bensaali, and A. Amira, Using big data and federated learning for generating energy efficiency recommendations, *Int. J.*

Data Sci. Anal., vol. 16, no. 3, pp. 353–369, 2023.

[57] G. Sumathi and J. Akilandeswari, Improved fuzzy weighted-iterative association rule based ontology postprocessing in data mining for query recommendation applications, *Comput. Intell.*, vol. 36, no. 2, pp. 773–782, 2020.

[58] H. Zhang, T. Huang, Z. Lv, S. Liu, and Z. Zhou, MCRS: A course recommendation system for MOOCs, *Multimed. Tools Appl.*, vol. 77, no. 6, pp. 7051–7069, 2018.

[59] S. H. Liao and C. A. Yang, Big data analytics of social network marketing and personalized recommendations, *Soc. Network Anal. Min.*, vol. 11, p. 21, 2021.

[60] P. K. Jain, G. Srivastava, J. C. W. Lin, and R. Pamula, Unscrambling customer recommendations: A novel LSTM ensemble approach in airline recommendation prediction using online reviews, *IEEE Trans. Comput. Soc. Sys.*, vol. 9, no. 6, pp. 1777–1784, 2022.

[61] R. A. Da Silva, A. M. De Paula Canuto, C. A. Da Silveira Barreto, and J. C. Xavier-Junior, Automatic recommendation method for classifier ensemble structure using meta-learning, *IEEE Access*, vol. 9, pp. 106254–106268, 2021.

[62] Q. Huang, Y. Xu, Y. Chen, H. Zhang, and F. Min, An adaptive mechanism for recommendation algorithm ensemble, *IEEE Access*, vol. 7, pp. 10331–10342, 2019.

[63] Z. Huang, G. Shan, J. Cheng, and J. Sun, TRec: An efficient recommendation system for hunting passengers with deep neural networks, *Neural Comput. Appl.*, vol. 31, no. Suppl1, pp. 209–222, 2019.

[64] G. Li, T. Zhu, J. Hua, T. Yuan, Z. Niu, T. Li, and H. Zhang, Asking images: Hybrid recommendation system for tourist spots by hierarchical sampling statistics and multimodal visual Bayesian personalized ranking, *IEEE Access*, vol. 7, pp. 126539–126560, 2019.

[65] X. Wang, C. Gao, J. Ding, Y. Li, and D. Jin, CMBPR: Category-aided multi-channel Bayesian personalized ranking for short video recommendation, *IEEE Access*, vol. 7, pp. 48209–48223, 2019.

[66] P. Lak, A. Bozanta, C. Kavaklioglu, M. Cevik, A. Basar, M. Petitclerc, and G. Wills, A replication study on implicit feedback recommender systems with application to the data visualization recommendation, *Expert Syst.*, vol. 39, no. 4, p. e12871, 2022.

[67] M. He, S. Zhang, and Q. Meng, Learning to style-aware Bayesian personalized ranking for visual recommendation, *IEEE Access*, vol. 7, pp. 14198–14205, 2019.

[68] H. Liu, Y. Tong, J. Han, P. Zhang, X. Lu, and H. Xiong, Incorporating multi-source urban data for personalized and context-aware multi-modal transportation recommendation, *IEEE Trans. Knowledge Data Eng.*, vol. 34, no. 2, pp. 723–735, 2022.

[69] A. Motwani, P. K. Shukla, and M. Pawar, Novel framework based on deep learning and cloud analytics for smart patient monitoring and recommendation (SPMR), *J. Ambient Intell. Human. Comput.*, vol. 14, no. 5, pp. 5565–5580, 2023.

[70] K. Gao, X. Yang, C. Wu, T. Qiao, X. Chen, M. Yang, and L. Chen, Exploiting location-based context for POI recommendation when traveling to a new region, *IEEE Access*, vol. 8, pp. 52404–52412, 2020.

[71] S. Y. Chou, J. S. R. Jang, and Y. H. Yang, Fast tensor factorization for large-scale context-aware recommendation from implicit feedback, *IEEE Trans. Big Data*, vol. 6, no. 1, pp. 201–208, 2020.

[72] M. Zhang, S. Wu, M. Gao, X. Jiang, K. Xu, and L. Wang, Personalized graph neural networks with attention mechanism for session-aware recommendation, *IEEE Trans. Knowledge Data Eng.*, vol. 34, no. 8, pp. 3946–3957, 2022.

[73] J. H. Chiang, C. Y. Ma, C. S. Wang, and P. Y. Hao, An adaptive, context-aware, and stacked attention network-based recommendation system to capture users' temporal preference, *IEEE Trans. Knowledge Data Eng.*, vol. 35, no. 4, pp. 3404–3418, 2023.

[74] W. Liu, Z. Lin, H. Zhu, J. Wang, and A. K. Sangaiah, Attention-based adaptive memory network for recommendation with review and rating, *IEEE Access*, vol. 8, pp. 113953–113966, 2020.

[75] D. Yin and S. Feng, Enhanced attention framework for multi-interest sequential recommendation, *IEEE Access*, vol. 10, pp. 67703–67712, 2022.

[76] X. Wang, Y. Sheng, H. Deng, and Z. Zhao, Top-N-targets-balanced recommendation based on attentional sequence-to-sequence learning, *IEEE Access*, vol. 7, pp. 120262–120272, 2019.

[77] Z. Cui, X. Sun, H. Chen, L. Pan, L. Cui, S. Liu, and G. Xu, Dynamic recommendation based on graph diffusion and Ebbinghaus curve, *IEEE Trans. Comput. Soc. Syst.*, vol. 11, no. 2, pp. 2755–2764, 2024.

[78] Z. Wang, Z. Wang, X. Li, Z. Yu, B. Guo, L. Chen, and X. Zhou, Exploring multi-dimension user-item interactions with attentional knowledge graph neural networks for recommendation, *IEEE Trans. Big Data*, vol. 9, no. 1, pp. 212–226, 2023.

[79] L. Sang, M. Xu, S. Qian, and X. Wu, Adversarial heterogeneous graph neural network for robust recommendation, *IEEE Trans. Comput. Soc. Syst.*, vol. 10, no. 5, pp. 2660–2671, 2023.

[80] R. Xie, Q. Liu, S. Liu, Z. Zhang, P. Cui, B. Zhang, and L. Lin, Improving accuracy and diversity in matching of recommendation with diversified preference network, *IEEE Trans. Big Data*, vol. 8, no. 4, pp. 955–967, 2022.

[81] I. U. Rehman, M. S. Hanif, Z. Ali, Z. Jan, C. B. Mawuli, and W. Ali, Empowering neural collaborative filtering with contextual features for multimedia recommendation, *Multimed. Syst.*, vol. 29, no. 4, pp. 2375–2388, 2023.

[82] L. Wu, X. He, X. Wang, K. Zhang, and M. Wang, A survey on accuracy-oriented neural recommendation:

From collaborative filtering to information-rich recommendation, *IEEE Trans. Knowledge Data Eng.*, vol. 35, no. 5, pp. 4425–4445, 2023.

[83] T. M. Selvi and V. Kavitha, A privacy-aware deep learning framework for health recommendation system on analysis of big data, *Vis. Comput.*, vol. 38, no. 2, pp. 385–403, 2022.

[84] B. Dong, Y. Zhu, L. Li, and X. Wu, Hybrid collaborative recommendation via dual-autoencoder, *IEEE Access*, vol. 8, pp. 46030–46040, 2020.

[85] M. D. Nguyen and Y. S. Cho, A variational autoencoder mixture model for online behavior recommendation, *IEEE Access*, vol. 8, pp. 132736–132747, 2020.

[86] A. Drif, H. E. Zerrad, and H. Cherifi, EnsVAE: Ensemble variational autoencoders for recommendations, *IEEE Access*, vol. 8, pp. 188335–188351, 2020.

[87] W. Zeng, J. Qin, and C. Wei, Neural collaborative autoencoder for recommendation with co-occurrence embedding, *IEEE Access*, vol. 9, pp. 163316–163324, 2021.

[88] C. Zhang, X. Wu, W. Yan, L. Wang, and L. Zhang, Attribute-aware graph recurrent networks for scholarly friend recommendation based on internet of scholars in scholarly big data, *IEEE Trans. Ind. Inf.*, vol. 16, no. 4, pp. 2707–2715, 2020.

[89] W. X. Zhao, S. Li, Y. He, E. Y. Chang, J. Wen, and X. Li, Connecting social media to e-commerce: Cold-start product recommendation using microblogging information, *IEEE Trans. Knowledge Data Eng.*, vol. 28, no. 5, pp. 1147–1159, 2016.

[90] L. Shen, Q. Liu, G. Chen, and S. Ji, Text-based price recommendation system for online rental houses, *Big Data Mining and Analytics*, vol. 3, no. 2, pp. 143–152, 2020.

[91] X. Chen, Z. Wang, H. Xu, J. Zhang, Y. Zhang, W. X. Zhao, and J. R. Wen, Data augmented sequential recommendation based on counterfactual thinking, *IEEE Trans. Knowledge Data Eng.*, vol. 35, no. 9, pp. 9181–9194, 2023.

[92] C. Xu, L. Xu, Y. Lu, H. Xu, and Z. Zhu, E-government recommendation algorithm based on probabilistic semantic cluster analysis in combination of improved collaborative filtering in big-data environment of government affairs, *Pers. Ubiquitous Comput.*, vol. 23, nos. 3&4, pp. 475–485, 2019.

[93] C. Cheng, Y. Lin, and C. Liao, Seq2CASE: Weakly supervised sequence to commentary aspect score estimation for recommendation, *IEEE Trans. Big Data*, vol. 9, no. 6, pp. 1670–1682, 2023.

[94] R. Rang, L. Xing, L. Zhang, H. Cai, and Z. Sun, Heterogeneous multi-behavior recommendation based on graph convolutional networks, *IEEE Access*, vol. 11, pp. 22574–22584, 2023.

[95] K. Liu, F. Xue, X. He, D. Guo, and R. Hong, Joint multi-grained popularity-aware graph convolution collaborative

filtering for recommendation, *IEEE Trans. Comput. Soc. Syst.*, vol. 10, no. 1, pp. 72–83, 2023.

[96] Z. Yang, Y. Wang, Y. Cheng, T. Zhang, and X. Wang, Recommendation model based on enhanced graph convolution that fuses review properties, *IEEE Trans. Comput. Soc. Syst.*, vol. 10, no. 5, pp. 2266–2278, 2023.

[97] X. Chen, J. Tian, X. Tian, and S. Liu, Fusing user reviews into heterogeneous information network recommendation model, *IEEE Access*, vol. 10, pp. 63672–63683, 2022.

[98] S. H. Ahammad, S. Dwarkanath, R. Joshi, B. T. P. Madhav, P. P. Priya, O. S. Faragallah, M. M. A. Eid, and A. N. Z. Rashed, Social media reviews based hotel recommendation system using collaborative filtering and big data, *Multimed. Tools Appl.*, vol. 83, no. 10, pp. 29569–29582, 2023.

[99] C. Wang, F. Feng, Y. Zhang, Q. Wang, X. Hu, and X. He, Rethinking missing data: Aleatoric uncertainty-aware recommendation, arXiv preprint arXiv: 2209.11679, 2023.

[100] X. Zhou, W. Liang, K. I. K. Wang, and L. T. Yang, Deep correlation mining based on hierarchical hybrid networks for heterogeneous big data recommendations, *IEEE Trans. Comput. Soc. Syst.*, vol. 8, no. 1, pp. 171–178, 2021.

[101] D. Wang, K. Liu, H. Xiong, and Y. Fu, Online POI recommendation: Learning dynamic geo-human interactions in streams, *IEEE Trans. Big Data*, vol. 9, no. 3, pp. 832–844, 2023.

[102] Q. Meng, H. Yan, B. Liu, X. Sun, M. Hu, and J. Cao, Recognize news transition from collective behavior for news recommendation, *ACM Trans. Inf. Syst.*, vol. 41, no. 4, p. 93, 2023.

[103] S. Sun, Y. Tang, Z. Dai, and F. Zhou, Self-attention network for session-based recommendation with streaming data input, *IEEE Access*, vol. 7, pp. 110499–110509, 2019.

[104] H. Wang, Y. Li, and F. Frimpong, Group recommendation via self-attention and collaborative metric learning model, *IEEE Access*, vol. 7, pp. 164844–164855, 2019.

[105] Z. Yang, L. Cui, X. Wang, T. Zhang, and Y. Cheng, MIAR: Interest-activated news recommendation by fusing multichannel information, *IEEE Trans. Comput. Soc. Syst.*, vol. 10, no. 6, pp. 3433–3443, 2023.

[106] Z. Cui, X. Sun, L. Pan, S. J. Liu, and G. Xu, Event-based incremental recommendation via factors mixed Hawkes process, *Inf. Sci.*, vol. 639, p. 119007, 2023.

[107] Q. Mei, Q. Hu, C. Yang, H. Zheng, and Z. Hu, Port recommendation system for alternative container port destinations using a novel neural language-based algorithm, *IEEE Access*, vol. 8, pp. 199970–199979, 2020.

[108] Z. Zhu and Y. Sun, Personalized information push system for education management based on big data

mode and collaborative filtering algorithm, *Soft Comput.*, vol. 27, no. 14, pp. 10057–10067, 2023.

[109] D. Xia, B. Shen, Y. Zheng, W. Zhang, D. Bai, Y. Hu, and H. Li, A bidirectional-a-star-based ant colony optimization algorithm for big-data-driven taxi route recommendation, *Multimed. Tools Appl.*, vol. 83, no. 6, pp. 16313–16335, 2023.

[110] M. Rostami, M. Oussalah, and V. Farrahi, A novel time-aware food recommender-system based on deep learning and graph clustering, *IEEE Access*, vol. 10, pp. 52508–52524, 2022.

[111] K. Sinaga and M. Yang, Unsupervised K-means clustering algorithm, *IEEE Access*, vol. 8, pp. 80716–80727, 2020.

[112] C. Morselli and C. Giguere, Legitimate strengths in criminal networks, *Crime, Law and Social Change*, vol. 45, no. 3, pp. 185–200, 2006.

[113] Recommendations: Netflix research, https://research.netflix.com/research-area/recommendations, 2024.

[114] S. Bouguezzi, How does the amazon recommendation system work? https://www.baeldung.com/cs/amazon-recommendation-system, 2024.

[115] P. Mangla, Spotify Music Recommendation Systems, PyimageSearch, https://pyimagesearch.com/2023/10/30/spotify-music-recommendation-systems/, 2024.

[116] B. Valarmathi and T. Chellatamilan, Classification of e-commerce online auctions and prediction of seller reputation, computer science, business, economics, https://repository.biust.ac.bw/bitstream/handle/123456789/158/65_Valarmathi%20B_Classification%20of%20e-commerce%20Online..pdf?sequence=1&isAllowed=y, 2024.

[117] N. Entezari, E. E. Papalexakis, H. Wang, S. Rao, and S. K. Prasad, Tensor-based complementary product recommendation, in *Proc. 2021 IEEE Int. Conf. Big Data (Big Data)*, Orlando, FL, USA, 2021, pp. 409–415.

[118] L. Wu, Browsemap: Collaborative filtering at LinkedIn, https://engineering.linkedin.com/recommender-systems/browsemap-collaborative-filtering-linkedin, 2024.

[119] V. Pavlik, What is the Google knowledge graph & how does it work? https://kenpeluso.me/what-is-the-google-knowledge-graph-how-does-it-work/, 2024.

[120] Guneet Lamba, Strategies for seasonal pricing for Airbnb for high and low-season, https://hello.pricelabs.co/seasonal-pricing-airbnb-for-high-and-low-season/, 2024.

[121] App rating prediction on Google store, https://medium.com/@insight_imi/app-rating-prediction-on-google-store-data-d6cde5050dba, 2024.

[122] The guide to YouTube SEO ranking, https://www.fatrank.com/ranking-youtube-videos-in-google-search/, 2024.

[123] Food discovery with Uber Eats: Recommending for the marketplace, https://www.uber.com/en-AE/blog/uber-eats-recommending-marketplace/, 2018.

[124] Y. Wang, Friend Recommendation using GraphSAGE, Stanford CS224W GraphML Tutorials, https://medium.com/stanford-cs224w/friend-recommendation-using-graphsage-ffcda2aaf8d6, 2024.

[125] P. Soni, Mastering recommendation engines with neural collaborative filtering, towards AI, https://pub.towardsai.net/mastering-recommendation-engines-with-neural-collaborative-filtering-059c4f11a470, 2024.

[126] S. Chakraborty, M. S. Hoque, N. R. Jeem, M. C. Biswas, D. Bardhan, and E. Lobaton, Fashion recommendation systems, models and methods: A review, *Informatics*, vol. 8, no. 3, p. 49, 2021.

[127] Ö. H. Cinar, Session-based recommender systems with recurrent neural networks, Medium 2023, https://medium.com/@omerhalidcinar/session-based-recommender-systems-with-recurrent-neural-networks-4c756dbdd7c9, 2024.

[128] K. Iman, Advances in next-track music recommendation (2017), https://ouci.dntb.gov.ua/en/works/98qnBA17/, 2017.

[129] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec, Graph convolutional neural networks for web-scale recommender systems, in *Proc. 24th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining*, London, UK, 2018, pp. 974–983.

[130] F. Daolio, Deep learning for fashion attributes, https://medium.com/asos-techblog/deep-learning-for-fashion-attributes-763c8c95034c, 2024.

[131] C. Yin and J. Han, Dynamic pricing model of e-commerce platforms based on deep reinforcement learning, *Comput. Model. Eng. Sci.*, vol. 127, no. 1, pp. 291–307, 2021.

[132] W. C. Kang and J. McAuley, Self-attentive sequential recommendation, in *Proc. 2018 IEEE Int. Conf. Data Mining (ICDM)*, Singapore, 2018, pp. 197–206.

[133] Market research solutions, how cluster analysis identifies market and customer segments, https://www.surveymonkey.com/market-research/resources/how-cluster-analysis-identifies-market-and-customer-segments/, 2024.

Kamal Taha has been an associate professor at Department of Computer Science, Khalifa University, United Arab Emirates (UAE) since 2010. He received the PhD degree in computer science from University of Texas at Arlington, USA. He has over 100 refereed publications that have appeared in prestigious top ranked journals, conference proceedings, and book chapters. Over 30 of his publications have appeared in IEEE transactions journals. He was an instructor of computer science at University of Texas at Arlington, USA from August 2008 to August 2010. He worked as an engineering specialist for Seagate Technology, USA from 1996 to 2005 (Seagate is a leading computer disc drive manufacturer in the US). His research interests span bioinformatics, information retrieval, data mining, databases, information forensics & security, and semiconductor wafers.

Chan Yeun received the MEng and PhD degrees in information security from University of London, UK in 1996 and 2000, respectively. Then he joined Toshiba TRL, Bristol, UK, and later became the vice president at Mobile Handset Research and Development Center, LG Electronics, Seoul, Republic of Korea in 2005. He was responsible for developing mobile TV technologies and related security. He left LG Electronics in 2007, and joined KAIST, Republic of Korea, until August 2008, and then Khalifa University, UAE in September 2008. He is currently a researcher of cybersecurity, including the IoT/USN security, cyber-physical system security, cloud/fog security, and cryptographic techniques; an associate professor at Department of Computer Science; and the cybersecurity leader of the Center for Cyber-Physical Systems (C2PS). He also enjoys lecturing for master cyber security and PhD engineering courses at Khalifa University. He has published more than 140 journal articles and conference papers, nine book chapters, and ten international patent applications. He also works on the editorial board of multiple international journals and on the steering committee of international conferences.

Paul D. Yoo is working at School of Computing and Mathematical Sciences, Birkbeck College, University of London, UK. Prior to this, he held academic/research posts in Cranfield (Defence Academy of the UK), Sydney, (USyd), and Republic of Korea (KAIST). In his career, he has amassed more than one hundred prestigious journal and conference publications, has been awarded more than $ 2.3 million in project funding, and a number of prestigious international and national awards for his work in advanced data analytics and machine learning research, notably IEEE Outstanding Leadership Award, Capital Markets CRC Award, Emirates Foundation Research Award, and the ICT Fund Award. Most recently, he won the prestigious Samsung award for research to protect IoT devices using machine-learning approach and Research England's Global Challenge Research Fund (GCRF) for research to protect global environment (e.g., marine resources) using edge intelligence techniques. Paul currently serves as an associate editor for *IEEE Transactions on Sustainable Computing*, *IEEE Access*, and *Journal of Big Data Research* (Elsevier).

Aya Taha is affiliated with Brighton College, Dubai, UAE. She has been actively involved in various research projects. She demonstrates exceptional skills in programming and is renowned for her leadership in initiating and leading numerous activities and volunteer efforts.