International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 12 | Dec - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

A Review on the Role of Generative AI in Personalization and Recommendation Systems

Mr. Subham Kumar Das[1], Mr. Ayush Sahoo[2], Dr. Goldi Soni[3]

[1]B. Tech Computer Science & Engineering, Amity University Chhattisgarh, Raipur, India- 493225
[2]B. Tech Computer Science & Engineering, Amity University Chhattisgarh, Raipur, India- 493225
[3]Assistant Professor, Amity University Chhattisgarh, Raipur, India- 493225

---***---

Abstract - Utilizing past behaviour data and forecasting future actions, recommender systems offer users individualized service support. Recommender systems have naturally incorporated artificial intelligence (AI), particularly computational intelligence and machine learning techniques and algorithms, to improve prediction accuracy and tackle problems with data sparsity and cold start. In-depth discussions of recommender system fundamentals and current practices are provided in this position paper, along with an analysis of how artificial intelligence (AI) might advance the field's technological advancements and applications. This study highlights fresh research paths and addresses current research difficulties in addition to reviewing recent theoretical and practical achievements. The work thoroughly examines a range of concerns pertaining to artificial intelligence (AI) recommender systems. Additionally, it assesses the progress made in these systems through the use of AI methods such as neural networks and deep learning, fuzzy approaches, transfer learning, genetic algorithms, evolutionary algorithms, and active learning.

***Key Words*:** Generative AI, Recommendation Systems, Content based recommendation systems, Reinforcement Learning, Deep Neural Network.

1. INTRODUCTION

Providing goods and services that specifically address the demands of each individual consumer can be difficult for firms in a cutthroat industry. Customer decision-making is made easier, and the user experience is improved with the aid of personalized e-services, which address a significant issue of information overload. Recommender systems were developed employing techniques and concepts from several artificial intelligence (AI) areas for user profile and preference finding when these customized e-services were first launched, twenty years ago.

The number of effective AI-driven apps has skyrocketed in the last several years. Successes include the self-driving car, The AI-powered program AlphaGo from Deep Mind, which famously beat a professional human player in the game of "Go," along with other developments in computer vision and speech recognition.

Recommender systems have a fantastic chance to capitalize on the remarkable advancements in artificial intelligence (AI), data analytics, and big data.

Lately, recommender systems have been incorporated with a variety of AI algorithms, which have improved customer happiness and the overall user experience. More accurate recommendations can be made thanks to AI than with traditional recommendation systems. A new age in recommender systems has been ushered in by this, offering more sophisticated data representations, generating deep insights into the links between people and objects, and unearthing extensive information in contextual, virtual, textural, and demographic data. The objective of this manuscript is to examine the latest and most innovative theoretical and practical advancements in the domain, to pinpoint constraints, and to suggest novel avenues for investigation concerning the advancement and utilization of artificial intelligence in recommender systems. It will make an effort to examine the problems associated with AI-powered recommender systems and the technology's potential to help comprehend massive data sets and turn data into knowledge. In this study, we have reviewed the advances in AI for recommender systems, including the use of fuzzy techniques, transfer learning, neural networks and deep learning, active learning, natural language processing, computer vision, and evolutionary computing.

The following are this paper's principal contributions:
1. An organized examination of some domains of artificial intelligence techniques and how they are used in recommender systems.
2. A summary of the models, techniques, and applications of cutting-edge AI in recommender systems.
3. An examination of open research problems that shed light on emerging trends and areas for further research, broadening the range of applications for AI methods in recommender systems.

2. LITERARY REVIEW

The paper explores the intersection of generative AI and recommendation systems, focusing on how emerging AI techniques enhance the ability to deliver personalized recommendations to users. Below is a review of its key sections:

Introduction and Context

The paper outlines the challenges businesses face in delivering tailored products and services to individual consumers. It highlights the evolution of recommendation systems from basic models to advanced AI-driven approaches. Generative AI, which includes neural networks, reinforcement learning, and transfer learning, has significantly advanced these systems, improving user satisfaction and the overall quality of recommendations.

Generative AI and Its Impact on Recommender Systems

The paper emphasizes that traditional recommender systems were mostly based on content-based and collaborative filtering techniques, but generative AI has introduced methods that offer greater accuracy. Deep neural networks (DNNs) have allowed these systems to better understand user behavior by analyzing large datasets, and reinforcement learning has enhanced the ability of systems to optimize long-term user engagement. The paper suggests that generative AI models, such as Generative Adversarial Networks (GANs), have played a crucial role in modelling complex user-item relationships.

AI Techniques in Personalization

Several AI methods have been integrated into recommendation systems, including:

- Deep Learning Models: These include convolutional neural networks (CNNs) for user review analysis and recurrent neural networks (RNNs) for sequential data, improving the predictive accuracy of recommendation models.
- Active Learning: This technique addresses data sparsity by prioritizing the most representative user-item interactions, which helps in refining recommendations.
- Reinforcement Learning: The paper discusses how reinforcement learning enables the system to balance exploration and exploitation, leading to more dynamic and personalized recommendations over time.

Challenges Addressed by AI in Recommender Systems

Generative AI has addressed some common challenges in traditional recommender systems:

- Cold Start Problem: AI models, especially deep learning, can help mitigate the cold start issue by analyzing minimal user data to generate relevant recommendations.
- Data Sparsity: Techniques like matrix factorization and deep learning-based models have been effective in managing the sparsity of user feedback, ensuring that systems can still make informed recommendations even with limited data.
- Long-Tail Problem: The paper points out that cross-domain recommender systems using generative AI techniques can increase the exposure of less popular items, addressing the issue of imbalanced data.

Future Directions

The paper also suggests several future research directions, particularly in the areas of:

- Privacy-Preserving Recommender Systems: As the use of AI in recommendation systems grows, ensuring user privacy is paramount. AI models that incorporate encryption and data perturbation techniques may offer solutions.
- Handling Imbalanced Data: The paper highlights the need for systems that can recommend "long-tail" items effectively, which are less popular but hold potential value for users and businesses alike.
- User Interaction and Explainability: With AI-driven personalization becoming more prevalent, enhancing transparency and understanding of how recommendations are generated will be key to user trust and system effectiveness.

3. RECOMENDER SYSTEMS: MAIN MODELS AND METHODS

The World Wide Web's tremendous rise in information and users now have many options thanks to the rapid development of e-services, which often makes decision-making more challenging.

The main purpose of recommender systems is to help those who lack expertise or understanding navigate the wide range of options available to them [1]. In order to forecast user preferences for items of interest, recommender systems leverage multiple information sources [2]. Over the past 20 years, recommender system research has garnered significant attention from both academia and industry. This research is often driven by the possible financial benefits that recommender systems might offer companies like Amazon [3]. Recommender systems were first used in e-commerce to alleviate the problem of information overload caused by Web 2.0.

They were then swiftly extended to the personalization of e-business, e-government, e-learning, and e-tourism [4]. Present-day online platforms like Amazon.com, YouTube, Netflix, Yahoo, Facebook, Last.fm, and Meetup are not complete without recommender systems. Recommender systems are designed to assess the usefulness and recommend ability of a thing, to put it succinctly. The core element of a recommender system is [5]:

$$f: U \times I \rightarrow D$$

This is a function to define the utility of a specific item i ∈ I to a user u ∈ U. D is the final recommendation list containing a set of items ranked according to the utility of all the items the user has not consumed. The utility of an item is presented in terms of user ratings. Recommender systems find an item for the user by maximizing the utility function, formulated as follows [5]:

$$\forall u \in U, \arg\max f(u,i). i \in I$$

Selecting a recommendation method affects how useful an item is predicted to be for a given user. Three types of recommendation systems can be distinguished by using the conventional taxonomies of earlier research [4-6]: content-based, collaborative filtering (CF)-based, and knowledge-based approaches. The ensuing subsections will examine these three groups.

4. CONTENT BASED RECOMENDER SYSTEMS ACKNOWLEDGEMENT

As the name implies, content-based recommender systems leverage a user's profile and the descriptive content of an item to anticipate its utility [7]. Content-based recommender systems aim to make recommendations for products that are similar to those that a certain user has previously expressed interest in.

Initial steps involve extracting various item attributes from descriptions and documents. A movie can be characterized, for example, by its genre, the director, the writer, the actors, the plot, etc. These attributes can be directly acquired from unstructured data, like news articles or articles, or from structured data, such tables. The vector space model with term frequency-inverse document frequency weighting is a keyword-based model that is one of the most often utilized retrieval strategies in content-based recommender systems [8]. Recommender systems that are based on content create a profile of a user's tastes based on the things in their consumption history. Usually, the profile contains information on the user's prior preferences. Because of this, the profiling process may be viewed as a traditional binary classification problem that has been extensively studied in the fields of machine learning and data mining. In this step, traditional techniques including decision trees, nearest neighbour algorithms, and Naïve Bayes are employed [9]. The algorithm determines the most pertinent items to create a suggestion list by comparing the item's properties with the user's profile once it has been created. Using the features gathered in the first two steps, the user profile and the item representation are filtered and matched to provide a suggestion from a content-based recommender system.

The result is to forward the matched items and remove those items the user tends to dislike, so the relevance evaluation of the recommendation is clearly dependent on the accuracy of the item's representation and the user's profile [10].

The content-based recommender system has several advantages [11, 12]. First, content-based recommendation is independent of the user since it is based on item representation. Consequently, the data sparsity issue does not affect this type of system.

Secondly, the new item cold-start issue is resolved by content-based recommender systems' ability to suggest new products to users. Lastly, a concise explanation of the recommendation outcome can be given by content-based recommender systems. In practical applications, this type of system's transparency is a huge benefit over other methods. Nevertheless, content-based recommender systems have a number of drawbacks [5, 13]. While these systems are able to solve the new item problem, they are still plagued by the new user problem since the recommendation result's accuracy is negatively impacted by the absence of user profile information. Moreover, users are always recommended similar goods by content-based systems, which causes the suggestion to become overly specialized. Because most users want to learn about new and stylish goods instead of being limited to items that are comparable to those they have already used, they tend to get bored with these types of recommendation lists. Another problem is that things aren't always easily represented in the precise way that content-based recommender systems need them to be. Therefore, articles or news items are a better fit for this type of system's recommendation than pictures or music.

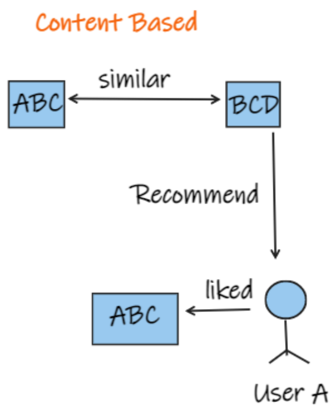

Fig 1: Content based recommendation system [source: towards ai]

5. COLABORATIVE FILTERING BASED RECOMENDER SYSTEMS

CF-based recommender systems deduce an item's utility based on other users' evaluations, as opposed to content-based recommender systems, which rely on a user's personal historical records but are not influenced by other users [13]. This method was swiftly implemented in the business more than 20 years ago [15] after extensive study in academia [14]. CF is still the most widely used method in recommender systems today [16]. Because the CF technique's basic premise is that users with similar interests would consume similar goods, a system that employs it depends on data supplied by users who share the same preferences as the given user. Predicting a user's ratings on unconsumed things from a user-item rating matrix is a classic issue in convolutional fields [17]. This problem is similar to the matrix completion problem. Memory-based CF and model-based CF are the two categories into which CF-based approaches are divided [18].

Memory-based CF is an early type of CF that may be further classified into two types: user-based CF and item-based CF [19]. Memory-based CF employs heuristic techniques to generate similarity values between individuals or items. This algorithm's ease of use, effectiveness, and precision in findings make it widely regarded. Memory-based CF is well known for its simple implementation, practical application, and relative effectiveness, although it still has some significant limitations [5]. Initially, it cannot resolve the issue of a cold start. The system does not have any ratings to use when a new user or item joins. Second, customers will give an item very few reviews if it is not new but is not well-liked by people. Memory-based CF has a restricted recommendation coverage because it is unlikely to suggest unpopular goods to consumers. Thirdly, it is unable to offer a recommendation in real time. When the user-item rating matrix has a high dimension, the heuristic procedure takes a while to produce a recommendation result. A pre-calculated and pre-stored weighting matrix in item-based CF can help to partially overcome this issue [19], but the scalability is still insufficient to fulfil real-world requirements.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 12 | Dec - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Instead of utilizing heuristic techniques, as was covered in the previous section, model-based CF uses using data mining or machine learning approaches to develop a model that forecasts a user's rating of items. Although this method was initially created to address the flaws in memory-based CF, it has been extensively researched to address issues in other fields. In addition to the user-item rating matrix, side information is used, such as location, tags and reviews [20]. If the rating matrix is paired with this auxiliary data, the model-based CF approach makes sense. As a result of the 2009 Netflix Prize competition, matrix factorization emerged as one of the most widely used algorithms in this domain [21]. The purpose of this is to make user space and item space comparable by projecting them onto the same latent factor space. Matrix factorization enjoys popularity for three key reasons. First, it significantly reduces the dimensions of the user-item rating matrix, ensuring the scalability of systems that utilize this technique. Second, the factorization process transforms a sparse rating matrix into a denser one, addressing the sparsity issue [22]. This enhancement allows users with limited ratings to receive more accurate recommendations, marking a notable advancement over memory-based approaches. Third, matrix factorization effectively accommodates various types of side information, which aids in profiling user preferences and enhances the overall performance of recommender systems [23].

6. KNOWLEDGE BASED RECOMMENDER SYSTEM

In knowledge-driven recommender systems, suggestions are derived from established knowledge or guidelines regarding user requirements and the functionalities of items [6]. In contrast to content-based and collaborative filtering (CF) approaches, knowledge-based recommender systems maintain a knowledge base that is developed from information gathered from a user's past interactions. This knowledge base encompasses prior issues, constraints, and the associated solutions. When the system faces a new recommendation challenge, it refers to the knowledge contained within this base [24]. Case-based reasoning employs earlier instances to address the current issue [25] and is a widely utilized method within knowledge-based systems. Finding similarities between products requires more structured representations than content-based recommender systems. This procedure entails contrasting the current scenario with a prior one.
one and adapting the solution accordingly. The use of knowledge-based recommendation techniques is especially beneficial in areas such as real estate sales, financial services, and healthcare decision support [26]. These sectors rely heavily on specialized domain expertise, and each situation is distinctly different. A key benefit of this approach is the absence of the new item/user dilemma, as existing knowledge is gathered and preserved within the knowledge base. Furthermore, users have the ability to set parameters for the recommendations provided [27]. Nonetheless, every advantage comes with its drawbacks; in this instance, the expenses associated with establishing and managing the system to build and sustain the knowledge base tend to be quite significant.

7. ARTIFICIAL INTELLIGENCE MAIN MODELS AND METHODS

Artificial intelligence is a rapidly evolving domain with applications that span from chess-playing programs to learning algorithms and disease diagnosis [28]. The primary aim of advancing AI techniques is to automate intelligent behaviours, which predominantly encompasses six key areas: knowledge engineering, reasoning, planning, communication, perception, and motion [29].
In detail, knowledge engineering involves methods for knowledge representation and modelling, allowing machines to comprehend and process information. Reasoning techniques are designed for tackling problems and making logical deductions. Planning assists machines in setting and achieving goals. Communication focuses on interpreting natural language and interacting with humans. Perception is concerned with analysing and processing inputs like images and speech. Finally, motion refers to manipulation and movement. Techniques from the first five domains, except motion, can greatly improve the creation of recommender systems to satisfy the wide range of information processing requirements.
As shown in Fig. 1, we will introduce eight main models and approaches in this section. With connections between them, deep neural networks, transfer learning, active learning, and fuzzy approaches are important symbols of knowledge and thinking. While computer vision focuses on visual recognition, natural language processing is clearly the main method for communication and perception, while evolutionary algorithms and reinforcement learning are linked to thinking and planning. Of these eight approaches, computer vision and natural language processing are two important areas where AI techniques are applied in recommender systems.

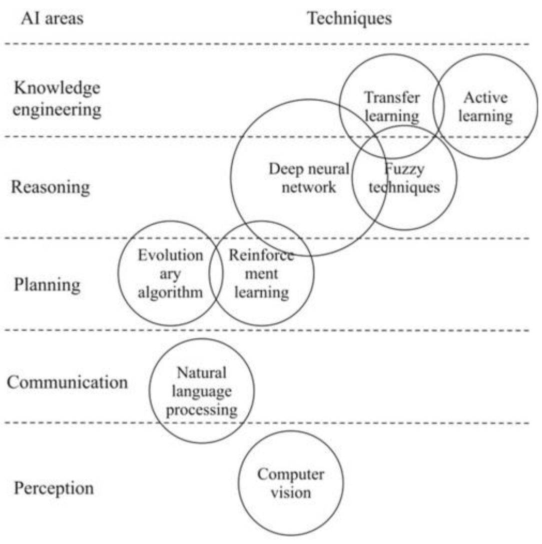

Fig.2: AI Areas and Techniques [source: google images]

7.1. Deep Neural Network

The network of neurons in the human brain serves as the model for neural networks. A group of neurons, also known as nodes, that receive and process signals from their interconnected counterparts make up a neural network. Depending on the incoming signal, each neuron can modify its internal state (activation).

allowing it to learn and modify activation weights and functions during the learning process. In the 1980s, neural networks fell out of favor and were largely disregarded by the machine learning community. However, by the late 1990s, a specific type of deep feedforward network known as convolutional neural networks (CNNs) was introduced, which proved to be significantly easier to train [30]. CNNs also demonstrated a greater ability to generalize compared to traditional neural networks, leading to their rapid adoption in fields such as speech recognition and computer vision [31]. Deep learning encompasses a wide range of different types [32].

Multilayer perceptrons (MLPs) [33] are feed-forward neural networks that comprise three or more layers equipped with non-linear activation functions. They enable the discovery of approximate solutions for various regression and classification tasks.

Autoencoders (AE) [34] are a type of unsupervised neural network designed to learn feature representations with the goals of dimensionality reduction, data compression, or noise reduction. The encoder, which processes the input data, and the decoder, which reconstructs the original input from the encoded representation, are the two primary parts of an autoencoder.

Convolutional neural networks (CNN) [35] are specialized for handling images and visual data. In addition to several hidden layers, which typically consist of convolutional, pooling, fully connected, and normalizing layers, they are organized with an input layer and an output layer.

Recurrent neural networks (RNNs) [36] are specifically engineered to handle sequential data, as their structure consists of node connections forming a directed graph. They utilize internal states to function as memory, enabling the retention of sequences over time. A notable example of RNNs is the long short-term memory (LSTM) network [37], which is particularly effective for time series forecasting.

Generative adversarial networks (GANs) [38] are employed for tasks that require unsupervised learning and consist of two competing models: a generative model and a discriminative model. These models work against each other to create samples that resemble the original data.

Graph neural networks (GNNs) [39] draw inspiration from convolutional neural networks (CNNs) and graph embedding techniques to represent the relationships among nodes within a graph structure, incorporating neighbourhood information. GNNs excel in tasks involving graph-structured data, such as representation learning, node classification, and link prediction, because of their high interpretability and outstanding performance.

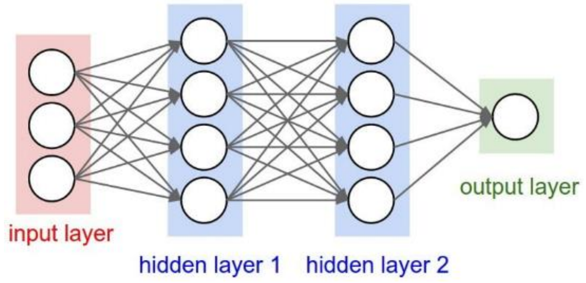

Fig 3: Deep Neural Networks [source: BMC Software, google images]

7.2. Reinforcement Learning

Reinforcement learning seeks to enhance reward outcomes through a series of actions performed by a learning agent in pursuit of a specific goal, where each action influences the subsequent situation or input in an interactive manner [40]. Unlike supervised learning, which depends on a labelled training dataset, reinforcement learning focuses on training an agent to operate in situations not included in the training data. Additionally, it differs Reinforcement learning seeks to accomplish long-term goals by interaction with the environment, in contrast to unsupervised learning, which finds patterns in unlabeled data.

The versatility of reinforcement learning contributes to its extensive application across various fields, including game theory [41], optimal control [42], swarm intelligence [43], as well as in domains such as healthcare [44] and psychology [45].

Reinforcement learning typically adheres to the principles of the Markov decision process [46] to illustrate how an agent interacts with its environment. The agent observes a state at each stage, decides on an action based on a policy, and is rewarded for that action. subsequently moving to the next step. A value function is employed to represent the long-term reward accumulated throughout the entire sequence of steps. A significant challenge in reinforcement learning is the trade-off between exploration and exploitation [47]. The agent must decide whether to adopt actions based on past experiences or to experiment with new actions that might yield greater rewards. Striking the right balance in this dilemma involves determining whether to take advantage of historically successful actions or to explore new strategies that ultimately lead to maximizing rewards. Reinforcement learning methods can be categorized based on value functions, policies, and models—differentiating into value-based or policy-based approaches, as well as off-policy or on-policy, model-based or model-free methods, and their hybrids [48]. Recently, the integration of deep neural networks with reinforcement learning has gained popularity, particularly exemplified by two notable successes: deep Q-networks [49] and AlphaGo [50]. The ability of reinforcement learning to handle high-dimensional states and actions has been greatly improved by the use of deep neural networks, making it an essential part of AI systems of the future.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 12 | Dec - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

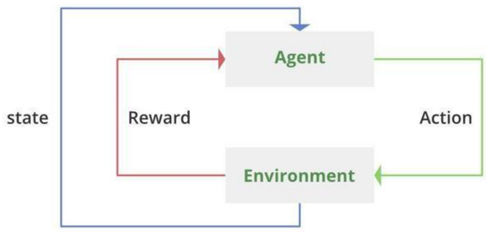

Fig 4: Reinforcement Learning [source: Geeks for Geeks]

8. RECOMENDER SYSTEMS WITH ARTIFICIAL INTELLIGENCE

A range of artificial intelligence algorithms have been created and employed in recommender systems to suit the growing need for recommendations driven by the surge of big data. Six AI techniques that have enhanced recommender systems will be discussed in this section.

8.1. Deep Neural Networks in recommender systems

Since the goal of recommendation is to rank items rather than classify them, neural networks are rarely used in recommender systems. To examine the ordinal nature of ratings, Salakhutdinov and associates first proposed a two-layer restricted Boltzmann machine (RBM). This approach garnered significant attention during the 2009 Netflix Prize competition [51], but subsequent research has been limited, with notable exceptions such as the work by Truyen et al., who expanded on the original study by examining RBM parameterization options in recommendation systems [52]. Conversely, deep learning has made remarkable strides in areas such as natural language processing, speech recognition, and computer vision [31]. The rise of abundant user-generated data, including comments and visual content related to items, has created a demand for systems that can integrate this information and recommend multimedia items, like images and videos. This need has led to the emergence of deep learning-based recommender systems [53]. In this section, we categorize these systems based on the various types of deep neural networks employed in recommendation tasks.

8.2. Convolutional Neural Network based recommender Systems

By combining two parallel neural networks, DeepCoNN models both users and items through their reviews [54]. The two convolutional neural networks (CNNs) are linked via a shared layer, which is enabled by factorization machines. To leverage user-contributed review information and tackle the issue of data sparsity, ConvMF incorporates CNNs into matrix factorization to enhance the accuracy of rating predictions [55]. Additionally, CNNs have been applied to the task of recommending hashtags in microblogging platforms by employing an attention mechanism during the hashtag selection process [56].

8.3. Active Learning in Recommender Systems

Every correlation between users and items within a recommender system, particularly those relying on explicit ratings or implicit user-item interactions, is vital for understanding user preferences and significantly impacts system performance. The issue of data sparsity in recommendations indicates that the more ratings collected from users, the better the system can perform in suggesting relevant items. However, obtaining ratings from users for all or even most items is tedious, labour-intensive, and nearly unfeasible. To address this, active learning has been implemented to assist recommender systems in identifying the most representative items and presenting them to users for ratings [64]. Active learning techniques have been used to improve the effectiveness and precision of recommender systems as user experience becomes more crucial and user involvement with computers is promoted in the current information era.

Early research employed active strategies that utilized pre-calculated boundaries on the value of information to minimize online computation time in recommender systems [65], although it was quickly recognized that the selection of items significantly affects rating predictions. Various active learning techniques have emerged, such as rating impact analysis and bootstrapping [66], which have been incorporated into popular recommendation models, including aspect models [67], decision trees [68], and matrix factorization [69]. Moreover, complex elements like ratings naturally given by users, the likelihood of users being able to provide a rating in response to system queries [70], the influence of specific items [71], and their attributes [72] have been integrated into the active learning framework. Additionally, these active learning strategies have been extended to multi-domain recommendation systems.

8.4. Reinforcement Learning in Recommender Systems

When recommender systems are used, the user and the system engage in an interactive process that is defined by several states and actions, aligning with principles of reinforcement learning. Unlike traditional recommender systems that primarily concentrate on predicting users' interests at a specific moment, reinforcement learning-based recommender systems focus on maximizing user engagement and satisfaction over the long term. In this reinforcement learning framework, the recommender system acts as a learning agent, user behaviors represent states, and the system-generated recommendations are considered actions. User feedback on these recommendations, such as click-through rates or time spent on a webpage, serves as the reward. The objective is to identify a policy or value function that enhances long-term rewards for users. A significant challenge in reinforcement learning arises from the vast number of available items, which creates a large action space for learning agents and complicates the system's dynamics.

Early studies primarily examined the balance between exploration and exploitation, often referred to as bandit problems [73]. One approach applied a Markov Decision Process (MDP) directly to recommender systems without addressing this balance [74], recommending the next item based on the previous k consumed items. Subsequent research incorporated the exploration-exploitation trade-off using linear reinforcement learning with theoretical backing [75]. Additionally, some studies have framed the interaction between users and recommender systems as multi-arm bandit problem [76], later expanded to include contextual information [77, 78].

The aforementioned research largely emphasizes immediate rewards, often neglecting long-term gains. Recently, deep reinforcement learning has attracted more attention due to advancements in deep Q-networks and deep deterministic policy gradients, which effectively mitigate challenges associated with both immediate and long-term rewards. Efforts to address the complexity of large and dynamic action spaces have been made using Actor-Critic architectures, which help reduce computational burdens. Negative user feedback is also being considered to enhance deep reinforcement learning-based recommendations through pair-wise regularization. A current trend in this field focuses on incorporating complex user behaviors and knowledge graph information to improve efficiency when dealing with extensive data and numerous items. The application of reinforcement learning techniques in industrial recommender systems is gaining traction, as seen in platforms like YouTube [79] and Alibaba [80]. The evolution of deep reinforcement learning-based recommender systems is expected to remain a vibrant area of research, increasingly influenced by practical industrial applications.

8.5. Computer Vision in Recommender Systems

Advances in computer vision technologies, especially in the areas of fashion analysis and visually-oriented products, have significantly improved recommender systems like clothing, jewelry, and images. The integration of image recognition with deep learning neural networks has led to remarkable outcomes in these systems.

A prominent application of this technology is in image recommendation. A dual-net deep network was introduced in [81], utilizing computer vision directly for image recommendation by correlating images with user preferences. Early research in other e-commerce recommendation fields leveraged features extracted from images through deep neural networks and combined them with established methods for clothing recommendations [82]. Subsequent studies have incorporated low-level features resembling elements of the human visual system, such as color attributes, into this framework [83]. Zhao et al. combined visual features derived from movie posters and still frames with a matrix factorization model to gain new insights into user preferences in movie recommendations [84]. Visual content has also played a role in point of interest recommendations, as user-uploaded images contain a wealth of landmarks [85]. To uncover shifting fashion trends among users, He et al. examined both non-visual and visual dimensions, incorporating temporal dynamics and deep convolutional networks [86]. Jaradat suggested the transfer of knowledge across domains by employing two convolutional neural networks—one for images and another for text—thereby tapping into user preferences concealed within social media platforms like Instagram [87].

Recommender systems must be equipped to profile users through multimedia data, with visual information being a key element. The application of multimodal fusion and multitask learning within recommender systems is essential for a comprehensive modeling of user preferences. Future fashion recommender systems are expected to indicate a strong demand for new functionalities like clothing design and coordination.

9. FUTURE DIRECTIONS

Recent advancements in recommender systems emphasize offering decision support by utilizing diverse information linked to item metadata, images, social networks, and reviews contributed by users. This paper examines the different domains of AI relevant to these systems and documents their progression. Considering that effective recommendations must align with user preferences while also enhancing the understanding of the interests of a wide array of users, we highlight several emerging research areas that could benefit from further investigation in the field of recommender systems.

10. LONG TAIL IN RECOMMENDER SYSTEMS(IMBALANCED DATA)

Long-tail items refer to those that are less popular and often overlooked by users. It is essential for recommender systems to give more attention to these items to assist users in discovering them. The reason long-tail items receive less attention is that there is limited data available about them, leading to their neglect by both users and e-commerce businesses. However, when appropriately utilized, long-tail items can provide significant advantages for both consumers and companies [88]. Cross-domain recommender systems have the potential to address the issue of long-tail items by leveraging knowledge from related yet distinct datasets, even when data is limited. As a result, creating recommender systems especially for long-tail commodities offers worthwhile research prospects.

11. SECURE AND PRIVACY PRESERVING RECOMMENDER SYSTEMS

The adoption of recommender systems has expanded significantly across various application domains, leading to increased user concerns regarding privacy. Consequently, users are often hesitant to share accurate information and preferences while using these systems, which negatively affects the performance of the recommender systems. Evolutionary algorithms are well-suited for developing recommender systems that protect privacy because of their capacity to handle several goals.

One approach involves employing encryption on user profiles, such as in a distributed collaborative filtering (CF) model that uses encrypted data [89]. However, this method raises concerns about high computational costs. Another strategy is to alter user profiles to mitigate the risk of data inference. For instance, study [90] demonstrates how adding randomness through data perturbation can protect privacy while maintaining recommendation accuracy. Research also investigates privacy preservation in collaborative filtering techniques, where similar users are grouped through data-independent hashing [91]. There is an urgent need to create secure and privacy-preserving recommender systems as more cross-platform technologies appear.

Furthermore, the use of recommender systems in high-risk privacy industries like banking and healthcare will propel the development of privacy-preserving methods.

12. CONCLUSION

In this position paper, we review eight domains of artificial intelligence, discuss how they are used in recommender systems, look at the unsolved research issues, and offer possible directions for further study on how to incorporate AI methods into these systems.

The paper emphasizes the ways in which AI methodologies can improve recommender systems and seeks to offer direction for both researchers and practitioners in this field.

REFERENCES

1. Shapira B, Ricci F, Kantor PB, Rokach L (2011) Recommendersystems handbook. Springer, New York
2. Bobadilla J, Ortega F, Hernando A, Gutiérrez A (2013) Recom-mender systems survey. Knowl Based Syst 46:109–132
3. Ben Schafer J, Konstan J, Riedl J (1999) Recommender systemsin e-commerce. In: Proceedings of the 1st ACM Conference on Electronic Commerce, 1999, pp 158–166
4. Lu J, Wu D, Mao M, Wang W, Zhang G (2015) Recommender system application developments: a survey. Decis Support Syst 74:12–32
5. Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possi- ble extensions. IEEE Trans Knowl Data Eng 17(6):734–749
6. Burke R (2002) Hybrid recommender systems: survey and exper-iments. User Model User-adapt Interact 12(4):331–370
7. Shardanand U, Maes P (1995) Social information filtering: algorithms for automating 'word of mouth'. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1995, pp 210–217
8. Salton G, Wong A, Yang C-S (1975) A vector space model for automatic indexing. Commun ACM 18(11):613–620
9. Sebastiani F (2002) Machine learning in automated text categorization. ACM Comput Surv 34(1):1–47
10. Herlocker JL, Konstan JA, Terveen LG, Riedl JT (2004) Evaluat- ing collaborative filtering recommender systems. ACM Trans InfSyst 22(1):5–53
11. Lops P, De Gemmis M, Semeraro G (2011) Content-based recommender systems: state of the art and trends. Recommender systems handbook. Springer, Berlin, pp 73–105
12. Shambour Q, Lu J (2012) A trust-semantic fusion-based recommendation approach for e-business applications. Decis Support Syst 54(1):768–780
13. Balabanovic´ M, Shoham Y (1997) Fab: content-based, collaborative recommendation. Commun ACM 40(3):66–72
14. Resnick P, Iacovou N, Suchak M, Bergstrom P, Riedl J (1994) GroupLens: an open architecture for collaborative filtering of net- news. In: Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, 1994, pp 175–186
15. Linden G, Smith B, York J (2003) Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Comput7(1):76–80
16. Liu H, Hu Z, Mian A, Tian H, Zhu X (2014) A new user similarity model to improve the accuracy of collaborative filtering. KnowlBased Syst 56:156–166
17. Hu Y, Zhang D, Ye J, Li X, He X (2013) Fast and accurate matrix completion via truncated nuclear norm regularization. IEEE Trans Pattern Anal Mach Intell 35(9):2117–2130
18. Su X, Khoshgoftaar TM (2009) A survey of collaborative filtering techniques. Adv Artif Intell 2009:1–19
19. Deshpande M, Karypis G (2004) Item-based top-n recommenda- tion algorithms. ACM Trans Inf Syst 22(1):143–177
20. Shi Y, Larson M, Hanjalic A (2014) Collaborative filtering beyondthe user-item matrix: a survey of the state of the art and future challenges. ACM Comput Surv 47(1):3
21. Koren Y, Bell R, Volinsky C (2009) Matrix factorization tech- niques for recommender systems. Computer 42(8):30–37
22. Luo X, Zhou M, Li S, You Z, Xia Y, Zhu Q (2016) A nonnegativelatent factor model for large-scale sparse matrices in recommender systems via alternating direction method. IEEE Trans Neural Netw Learn Syst 27(3):579–592
23. Liu B, Xiong H, Papadimitriou S, Fu Y, Yao Z (2015) A gen- eral geographical probabilistic factor model for point of interest recommendation. IEEE Trans Knowl Data Eng 27(5):1167–1179.
24. Smyth B (2007) Case-based recommendation. The adaptive web. Springer, Berlin, pp 342–376
25. Aamodt A, Plaza E (1994) Case-based reasoning: foundational issues, methodological variations, and system approaches. AI Commun 7(1):39–59
26. Felfernig A, Friedrich G, Jannach D, Zanker M (2011) Devel- oping constraint-based recommenders. Recommender systems handbook. Springer, Berlin, pp 187–215
27. Felfernig A, Burke R (2008) Constraint-based recommender sys- tems: technologies and research issues. In: Proceedings of the 10th International Conference on Electronic Commerce, 2008, p3
28. Luger GF (2005) Artificial intelligence: structures and strategies for complex problem solving. Pearson Education, London
29. Russell SJ, Norvig P (2016) Artificial intelligence: a modern approach. Pearson Education Limited, Malaysia
30. LeCun Y, Bengio Y (1995) Convolutional networks for images,speech, and time series. Handb Brain Theor Neural Netw 3361(10):1995
31. Lecun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
32. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT press, Cambridge
33. Murtagh F (1991) Multilayer perceptrons for classification and regression. Neurocomputing 2(5–6):183–197

34. Wang Y, Yao H, Zhao S (2016) Auto-encoder based dimension-ality reduction. Neurocomputing 184:232–242
35. Lawrence S, Giles CL, Tsoi AC, Back AD (1997) Face recog- nition: a convolutional neural-network approach. IEEE Trans Neural Netw 8(1):98–113
36. Rumelhart DE, Hinton GE, Williams RJ (1986) Learn- ing representations by back-propagating errors. Nature 323(6088):533–536
37. Hochreiter S, Schmidhuber J (1997) LSTM can solve hard long time lag problems. Advances in neural information processing systems. MIT Press, Cambridge, pp 473–479
38. Goodfellow I et al (2014) Generative adversarial nets. Advances in neural information processing systems. MIT Press, Cambridge, pp 2672–2680
39. Zhou J et al (2018) Graph neural networks: a review of methodsand applications. arXiv Prepr. arXiv1812.08434.
40. Sutton RS, Barto AG (2011) Reinforcement learning: An intro- duction. MIT Press, Cambridge
41. Peng et al P (2017) Multiagent bidirectionally-coordinated nets:emergence of human-level coordination in learning to play star- craft combat games. arXiv Prepr. arXiv1703.10069
42. Bai W, Li T, Tong S (2020) NN reinforcement learning adaptivecontrol for a class of nonstrict-feedback discrete-time systems. IEEE Trans Cybern
43. Hüttenrauch M, Adrian S, Neumann G (2019) Deep reinforcement learning for swarm systems. J Mach Learn Res 20(54):1–31
44. Neftci EO, Averbeck BB (2019) Reinforcement learning in arti-ficial and biological systems. Nat Mach Intell 1(3):133–143
45. Botvinick M, Ritter S, Wang JX, Kurth-Nelson Z, Blundell C, Hassabis D (2019) Reinforcement learning, fast and slow. Trends Cogn Sci 23(5):408–422
46. Bellman R (1957) A Markovian decision process. J Math Mech 679–684
47. Henderson P, Islam R, Bachman P, Pineau J, Precup D, Meger D (2017) Deep reinforcement learning that matters. arXiv Prepr. arXiv1709.06560
48. Kaelbling LP, Littman ML, Moore AW (1996) Reinforcement learning: a survey. J Artif Intell Res 4:237–285
49. Mnih V et al (2015) Human-level control through deep reinforce- ment learning. Nature 518(7540):529–533
50. Silver D et al (2016) Mastering the game of Go with deep neuralnetworks and tree search. Nature 529(7587):484–489
51. Salakhutdinov R, Mnih A, Hinton G (2007) Restricted Boltzmann machines for collaborative filtering. In: Proceedings of the 24th International Conference on Machine Learning, 2007, pp 791–798
52. Truyen TT, Phung DQ, Venkatesh S (2009) Ordinal Boltzmann machines for collaborative filtering. In: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, 2009, pp 548–556
53. Zhang S, Yao L (2017) Deep learning based recommender system: a survey and new perspectives. ACM J Comput Cult Herit Artic1(35):1–35
54. Diao Q, Qiu M, Wu CY, Smola AJ, Jiang J, Wang C (2014) Jointlymodeling aspects, ratings and sentiments for movie recommen- dation. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014, pp 193–202
55. Kim D, Park C, Oh J, Lee S, Yu H (2016) Convolutional matrix factorization for document context-aware recommendation. In: RecSys 2016—Proceedings of the 10th ACM Conference on Rec- ommender Systems, 2016, pp 233–240
56. Yuyun G, Qi Z (2016) Hashtag recommendation using attention- based convolutional neural network. In: International Joint Con-ference on Artificial Intelligence, 2016, pp 2782–2788
57. Wang J et al. (2017) IRGAN: a minimax game for unifying generative and discriminative information retrieval models. In: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017, pp 515–524
58. He X, He Z, Du X, Chua TS (2018) Adversarial personalized rank-ing for recommendation. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018, pp 355–364
59. Yang D, Guo Z, Wang Z, Jiang J, Xiao Y, Wang W (2018) A knowledge-enhanced deep recommendation framework incorpo- rating GAN-based models. In: 2018 IEEE International Confer- ence on Data Mining, 2018, pp 1368–1373
60. Tang J, Du X, He X, Yuan F, Tian Q, Chua T-S (2019) Adversarial training towards robust multimedia recommender system. IEEE Trans Knowl Data Eng 32(5):855–867
61. Ying R, He R, Chen K, Eksombatchai P, Hamilton WL, LeskovecJ (2018) Graph convolutional neural networks for web-scale rec-ommender systems. In: Proceedings of the 24th ACM SIGKDDInternational Conference on Knowledge Discovery and Data Min- ing, 2018, pp 974–983
62. Yin R, Li K, Zhang G, Lu J (2019) A deeper graph neural network for recommender systems. Knowl Based Syst 185:105020
63. Wu S, Tang Y, Zhu Y, Wang L, Xie X, Tan T (2019) Session-based recommendation with graph neural Networks. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2019, vol. 33, pp 346–353
64. Elahi M, Ricci F, Rubens N (2016) A survey of active learning in collaborative filtering recommender systems. Comput Sci Rev20:29–50
65. Boutilier C, Zemel RS, Marlin B (2003) Active collaborative fil- tering. In: Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence, 2003, pp 98–106
66. Mello CE, Aufaure MA, Zimbrao G (2010) Active learning driven by rating impact analysis. In: Proceedings of the 4th ACM Con- ference on Recommender Systems, 2010, pp 341–344
67. Golbandi N, Koren Y, Lempel R (2010) On bootstrapping recom- mender systems. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, 2010,pp 1805–1808
68. Karimi R, Freudenthaler C, Nanopoulos A, Schmidt-Thieme L (2011) Active learning for aspect model in recommender systems. In: IEEE Symposium on Computational Intelligence and Data Mining, 2011, pp 162–167
69. Golbandi N, Koren Y, Lempel R (2011) Adaptive bootstrappingof recommender systems using decision trees. In: Proceedings of the 4th ACM International Conference on Web Search and DataMining, 2011, pp 595–604
70. Karimi R, Freudenthaler C, Nanopoulos A, Schmidt-Thieme L (2011) Non-myopic active learning for recommender systems based on matrix factorization. In: IEEE International Conference on Information Reuse & Integration, 2011, pp 299–303
71. Wiesner M, Pfeifer D (2010) Adapting recommender systems to the requirements of personal health record systems. In: Proceed- ings of the 1st ACM International Health Informatics Symposium, 2010, pp 410–414
72. Elahi M, Ricci F, Rubens N (2012) Adapting to natural rating acquisition with combined active learning strategies. In: Inter- national Symposium on Methodologies for Intelligent Systems, 2012, pp 254–263

73. He L, Liu NN, Yang Q (2011) Active dual collaborative filter- ing with both item and attribute feedback. In: Proceedings of the National Conference on Artificial Intelligence, 2011, vol. 2, pp 1186–1191
74. Berry DA, Fristedt B (1985) Bandit problems: sequential allo- cation of experiments (Monographs on statistics and applied probability). London Chapman Hall 5(71–87):7
75. Shani G, Heckerman D, Brafman RI (2005) An MDP-based rec-ommender system. J Mach Learn Res 6:1265–1295
76. Warlop R et al (2018) Fighting boredom in recommender systems with linear reinforcement learning. No. NeurIPS, 2018
77. Wang H, Wu Q, Wang H (2017) Factorization bandits for inter-active recommendation. AAAI 17:2695–2702
78. Li L, Chu W, Langford J, Schapire RE (2010) A contextual-bandit approach to personalized news article recommendation. In: Proc. 19th Int. Conf. World Wide Web, pp. 661–670, 2010
79. Zeng C, Wang Q, Mokhtari S, Li T (2016) Online context-aware recommendation with time varying multi-armed bandit. In: Pro- ceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp 2025–2034
80. Ie E et al (2019) SLateq: a tractable decomposition for reinforce- ment learning with recommendation sets. In: Int Jt Conf Artif Intell, vol. 2019-Augus, pp 2592–2599, 2019
81. Lei C, Liu D, Li W, Zha ZJ, Li H (2016) Comparative deep learning of hybrid representations for image recommendations.
82. He R, McAuley J (2015) VBPR: visual bayesian personalized ranking from implicit feedback. In: AAAI, 2015, pp 144–150
83. Gaspar P (2017) User preferences analysis using visual stimuli. In: Proceedings of the 11th ACM Conference on RecommenderSystems, 2017, pp 436–440
84. Zhao L, Lu Z, Pan SJ, Yang Q (2016) Matrix factorization+ for movie recommendation. In: International Joint Conference on Artificial Intelligence, 2016, pp 3945–3951
85. Wang S, Wang Y, Tang J, Shu K, Ranganath S, Liu H (2017) Whatyourimagesreveal: exploitingvisualcontentsforpoint-of-interest recommendation. In: Proceedings of the 26th International Con-ference on World Wide Web, 2017, pp 391–400
86. He R, McAuley J (2016) Ups and downs: modeling the visual evolution of fashion trends with one-class collaborative filtering. In: Proceedings of the 25th International Conference on World Wide Web, 2016, pp 507–517
87. Jaradat S (2017) Deep cross-domain fashion recommendation. In: Proceedings of the 11th ACM Conference on RecommenderSystems, 2017, pp 407–410
88. Yin H, Cui B, Li J, Yao J, Chen C (2012) Challenging the long tail recommendation. In: Proceedings of the VLDB Endowment, 2012, vol 5, no 9, pp 896–907
89. Canny J (2002) Collaborative filtering with privacy. In: Proc. IEEESymp. Secur. Priv., vol. 2002-Jan, pp 45–57, 2002
90. Kikuchi H, Mochizuki A (2013) Privacy-preserving collaborative filtering using randomized response. J Inf Process 21(4):617–623
91. Chow R, Pathak MA, Wang C (2012) A practical system for privacy-preserving collaborative filtering. In: Proc. 12th IEEE Int. Conf. Data Min. Work. ICDMW 2012, pp 547–554, 2012
92. Bostandjiev S, O'Donovan J, Höllerer T (2012) TasteWeights: a visual interactive hybrid recommender system. In: Proceedings of the sixth ACM conference on Recommender systems, 2012, pp 35–42
93. Wang W, Zhang G, Lu J (2017) Hierarchy visualization for group recommender systems. In: IEEE Trans Syst Man Cybern Syst, pp 1–12, 2017
94. Hernando A, Moya R, Ortega F, Bobadilla J (2014) Hierarchi- cal graph maps for visualization of collaborative recommender systems. J Inf Sci 40(1):97–106