International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 5
Sep-Oct-2026, PP 01-06

Ai-Based Recommendation Systems: A Review Of Collaborative, Content-Based, Deep And Sequential Approaches For Personalization

Kachale Manali Bhagwan[1] Prof. C. A. Kolse[2] Prof. M. C. Kolse[3] Prof. S. K. Said[4]
Department of Artificial Intelligence and Data Science
Jaihind College of Engineering
Pune, Maharashtra, India

Abstract—Recommendation systems deliver personalized experiences on e-commerce, media streaming and social platforms by suggesting items that users are likely to value. Traditional collaborative filtering and content-based filtering face data sparsity and the cold-start problem for new users and items. This paper reviews AI-based recommendation with machine learning and deep learning. It explains neighbourhood-based and matrix-factorization collaborative filtering, content-based filtering and hybrid designs, then surveys deep learning recommenders including neural collaborative filtering, wide-and-deep and factorization-machine models for click-through prediction, two-stage candidate generation and ranking at industrial scale, and self-attentive and bidirectional transformer models that learn from sequences of user actions. Evaluation with benchmark datasets and ranking metrics, and strategies for cold start, are discussed. A five-layer real-time recommendation architecture comprising data collection, feature processing, candidate generation, ranking and serving with feedback is proposed, along with research gaps, objectives, a comparison of approaches and challenges including popularity bias, filter bubbles, privacy and explainability. The review concludes that hybrid deep recommenders that combine behaviour, content and context deliver accurate real-time suggestions that improve user engagement.

Index Terms—Recommendation systems, collaborative filtering, content-based filtering, deep learning, sequential recommendation, cold start, personalization

I. Introduction

An online store may sell millions of products, and a streaming service may host tens of thousands of titles. Without help, users cannot find what they want, and businesses lose sales and engagement. Recommendation systems solve this by predicting which items each user is likely to enjoy.

Early systems used simple rules such as “customers who bought this also bought”. Sarwar *et al.* [1] formalized item-based collaborative filtering, and Koren *et al.* [2] popularized matrix factorization. Deep learning has since enabled models that combine many signals and learn from sequences of behaviour. Zhang *et al.* [3] surveyed deep learning-based recommendation and its new perspectives.

This paper reviews these approaches and proposes a real-time recommendation architecture.

II. Related Work

Sarwar *et al.* [1] showed that computing similarities between items rather than users gives better scalability and quality. Koren *et al.* [2] described matrix factorization methods that won the Netflix Prize by representing users and items as latent vectors.

He *et al.* [4] proposed neural collaborative filtering, replacing the dot product with a neural network. Cheng *et al.* [5] introduced wide-and-deep learning at Google Play, combining a memorizing linear model with a generalizing deep network, and Guo *et al.* [6] proposed DeepFM, which learns low- and high-order feature interactions without manual feature engineering.

Covington *et al.* [7] described YouTube's two-stage deep recommender: a candidate-generation network selects hundreds of videos from millions, and a ranking network orders them. For sequences, Kang and McAuley [8] applied self-attention to users' action histories in SASRec, and Sun *et al.* [9] proposed BERT4Rec with bidirectional transformers. Schein *et al.* [10] studied methods and metrics for cold-start recommendation, and Harper and Konstan [11] documented the MovieLens datasets widely used for evaluation. Table I summarizes these works.

TABLE I Summary of Literature Survey

Ref.	Authors	Year	Technique Used	Merits	Demerits
[1]	Sarwar et al.	2001	Item-based collaborative filtering	Scalable, accurate similarity	Sparse data limits
[2]	Koren et al.	2009	Matrix factorization	Strong latent-factor accuracy	Cold start
[5]	Cheng et al.	2016	Wide & Deep	Memorization + generalization	Manual cross features
[7]	Covington et al.	2016	Two-stage deep recommender	Scaled to millions of videos	Complex infrastructure
[4]	He et al.	2017	Neural collaborative filtering	Learned interaction function	Implicit feedback only
[6]	Guo et al.	2017	DeepFM	Automatic feature interactions	CTR task focus
[8]	Kang and McAuley	2018	Self-attentive sequential model	Captures evolving interests	Long sequences costly
[9]	Sun et al.	2019	BERT4Rec	Bidirectional sequence modelling	Training cost

III. Preliminaries

3.1 Recommendation Approaches

Fig. 1 classifies approaches: collaborative filtering uses patterns of user behaviour, content-based filtering uses item attributes [12], hybrid methods combine both, and deep and sequential models learn complex patterns.

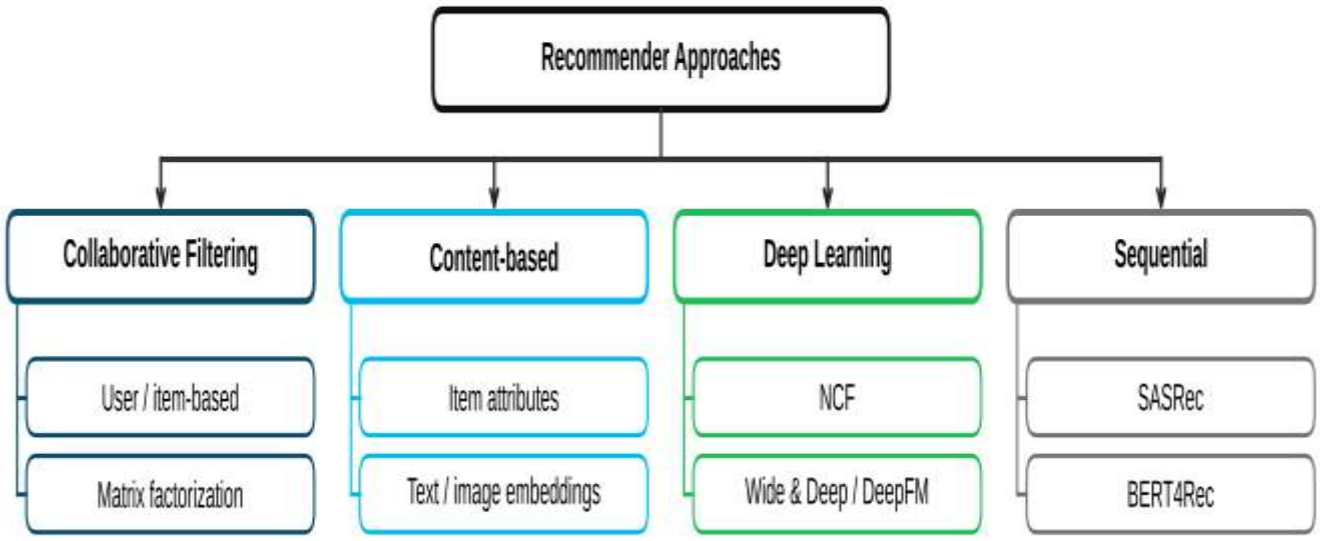


Fig. 1. Classification of recommendation approaches.

3.2 Explicit and Implicit Feedback

Explicit feedback includes ratings and reviews; implicit feedback includes clicks, views, purchases and watch time. Implicit feedback is abundant but noisy, and models such as Bayesian personalized ranking [13] learn from it by ranking observed items above unobserved ones.

IV. Problem Statement and Objectives

4.1 Identified Research Gaps

1. 1) New users and items still receive poor recommendations because of the cold-start problem.
2. 2) Popularity bias concentrates exposure on already popular items.
3. 3) Offline metrics often fail to predict online user satisfaction.
4. 4) Explanations for recommendations are rarely provided.
5. 5) Collecting detailed behaviour raises privacy concerns.

4.2 Objectives of the Review

- To review classical, deep and sequential recommendation approaches.
- To examine solutions for sparsity and cold start.
- To propose a real-time, layered recommendation architecture.
- To compare approaches in accuracy, scalability and cold-start handling.
- To identify challenges and future directions.

4.3 Proposed Framework

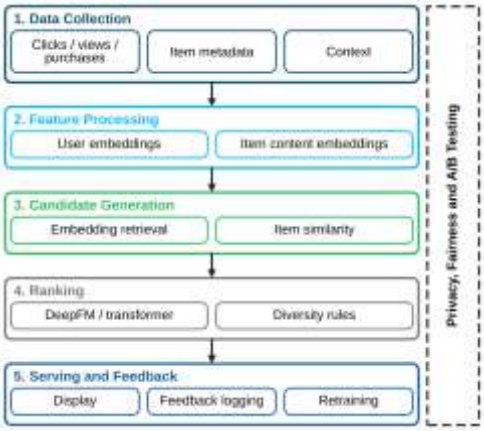


Fig. 2 shows the proposed architecture.

Fig. 2. Proposed real-time recommendation architecture.

4.4 Data Collection Layer

User interactions, item metadata and context such as time, device and location are logged in real time.

B. Feature Processing Layer

Features include user history embeddings, item text and image embeddings and contextual signals; content features help with new items.

4.5 Candidate Generation Layer

Embedding-based retrieval and item-to-item similarity quickly select a few hundred candidates from the full catalogue.

4.6 Ranking Layer

A deep ranking model such as DeepFM or a sequential transformer scores candidates; business rules ensure diversity and freshness.

4.7 Serving and Feedback Layer

Ranked items are displayed, and user responses are fed back for continuous learning and A/B testing.

V. Working Methodology

Fig. 3 describes the workflow. In matrix factorization the predicted preference of user u for item i is

$$\hat{r_{ui}} = \mu + b_u + b_i + p_u{}^T q_i \quad (1)$$

where μ is the global mean, b_u and b_i are user and item biases and p_u, q_i are latent vectors. Ranking quality is measured with normalized discounted cumulative gain:

$$NDCG@K = DCG@K / IDCG@K \quad (2)$$

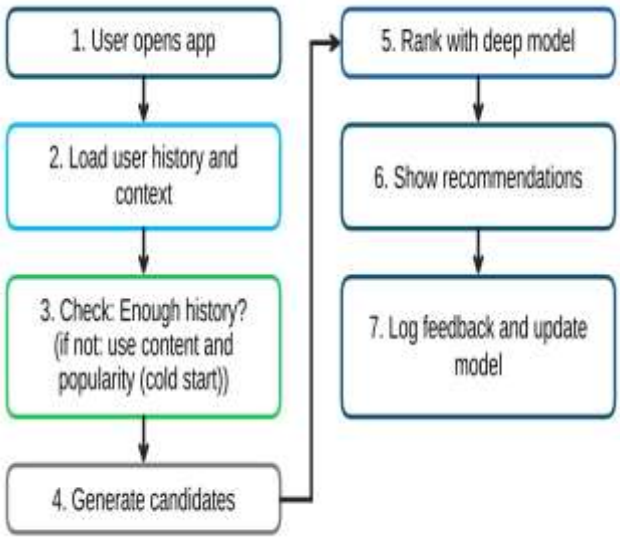


Fig. 3. Recommendation workflow.

VI. Discussion

6.1 Comparison of Approaches

Table II compares approaches. Collaborative filtering is effective with rich behaviour data, content-based methods handle new items, deep hybrid models combine many signals, and sequential models capture changing interests.

Approach	Accuracy	Cold Start	Scalability
User / item CF	Medium	Poor	Medium
Matrix factorization	High	Poor	High
Content-based	Medium	Good (items)	High
Deep hybrid (DeepFM)	Very high	Medium	High
Sequential transformer	Very high	Medium	Medium

TABLE II Comparison of Recommendation Approaches

6.2 Open Challenges and Future Work

Table III summarizes challenges. Future directions include conversational recommenders powered by language models, fairness-aware ranking, privacy-preserving federated recommendation and explanations that tell users why an item was suggested.

Challenge	Mitigation
Cold start	Content features; onboarding questions
Popularity bias	Re-ranking for diversity
Filter bubbles	Exploration; serendipity
Privacy	Federated and on-device learning
Explainability	Reason labels for recommendations

TABLE III Challenges and Mitigation

VII. Conclusion and Future Work

AI-based recommendation systems turn large catalogues into personalized experiences. The reviewed literature shows the progression from item similarity and matrix factorization to deep hybrid and sequential models that power real-time recommendations at scale. A layered architecture with candidate generation, deep ranking and continuous feedback can provide accurate, timely suggestions, while attention to cold start, bias, privacy and explainability ensures they remain fair and trustworthy.

7.1 Acknowledgment

The authors express their sincere gratitude to the Department of Artificial Intelligence and Data Science, Jaihind College of Engineering, Pune, for providing the facilities, guidance and encouragement needed to carry out this work. The authors also thank the faculty members and colleagues whose valuable suggestions helped improve the quality of this review.

References

1. B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Item-based collaborative filtering recommendation algorithms,” in Proc. 10th Int. Conf. World Wide Web (WWW), 2001, pp. 285–295.
2. Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” Computer, vol. 42, no. 8, pp. 30–37, 2009.
3. S. Zhang, L. Yao, A. Sun, and Y. Tay, “Deep learning based recommender system: A survey and new perspectives,” ACM Comput. Surv., vol. 52, no. 1, Art. no. 5, 2019.
4. X. He et al., “Neural collaborative filtering,” in Proc. 26th Int. Conf. World Wide Web (WWW), 2017, pp. 173–182.
5. H.-T. Cheng et al., “Wide & deep learning for recommender systems,” in Proc. 1st Workshop Deep Learn. Recommender Syst. (DLRS), 2016, pp. 7–10.

6. H. Guo, R. Tang, Y. Ye, Z. Li, and X. He, “DeepFM: A factorization-machine based neural network for CTR prediction,” in Proc. 26th IJCAI, 2017, pp. 1725–1731.
7. P. Covington, J. Adams, and E. Sargin, “Deep neural networks for YouTube recommendations,” in Proc. 10th ACM Conf. Recommender Syst. (RecSys), 2016, pp. 191–198.
8. W.-C. Kang and J. McAuley, “Self-attentive sequential recommendation,” in Proc. IEEE Int. Conf. Data Min. (ICDM), 2018, pp. 197–206.
9. F. Sun et al., “BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer,” in Proc. 28th ACM CIKM, 2019, pp. 1441–1450.
10. A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, “Methods and metrics for cold-start recommendations,” in Proc. 25th ACM SIGIR, 2002, pp. 253–260.
11. F. M. Harper and J. A. Konstan, “The MovieLens datasets: History and context,” ACM Trans. Interact. Intell. Syst., vol. 5, no. 4, Art. no. 19, 2015.
12. P. Lops, M. de Gemmis, and G. Semeraro, “Content-based recommender systems: State of the art and trends,” in Recommender Systems Handbook. Boston, MA, USA: Springer, 2011, pp. 73–105.
13. S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, “BPR: Bayesian personalized ranking from implicit feedback,” in Proc. 25th Conf. Uncertainty Artif. Intell. (UAI), 2009, pp. 452–461.