International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Beyond Ratings: A Context-Aware Movie Recommender with Knowledge Graphs and Confidence-Guided Ranking

Oshin Verma[1], Archana Kumar[2]

[1]Scholar, Department of AI-DS, ADGIPS, GGSIPU, Delhi
[2]Supervisor, Department of AI-DS, ADGIPS, GGSIPU, Delhi

Abstract
Movie recommendation must balance relevance with sparse feedback, new-user and new-item cold start, limited explanations, and changing interests. This paper synthesizes eleven supplied studies on hybrid filtering, sentiment, facial emotion, scalable collaborative filtering, and knowledge-graph recommendation, then proposes a modular confidence-aware framework. The design generates candidates from item metadata, graph relations, and collaborative signals; combines them through a cascade; and applies context only when the signal is available, sufficiently reliable, and explicitly permitted. A fallback path supports users with little history, while explanations identify the principal metadata or graph relations behind a recommendation. The paper specifies the architecture, scoring logic, privacy boundaries, and a reproducible evaluation plan. It reports no new experiment: the numerical results discussed are those reported by individual source studies and are not treated as directly comparable because their datasets and protocols differ. The proposed framework therefore serves as a testable design rather than a validated performance claim.

Keywords: Movie Recommendation, Hybrid Filtering, Knowledge Graph, Context-Aware Ranking, Sentiment Analysis, Privacy

1 Introduction

Streaming catalogues make it easy to access films but difficult to choose among them. A recommender reduces search effort by estimating which items may be useful to a particular viewer. Yet any one source of evidence is incomplete. Content-based filtering can retrieve films that resemble a viewer's known interests, but tends to repeat familiar attributes. Collaborative filtering can surface unexpected items from crowd behavior, but requires interaction overlap and may fail for new users or titles. Popularity rankings help when little is known, but are weakly personalized [1, 2, 3].

Recent movie-recommendation work explores combinations of these signals. Hybrid approaches have combined content features with user or item collaboration, while sentiment analysis attempts to recover preference evidence from reviews or social posts [4, 5, 6]. Knowledge graphs add explicit relations among films, people, genres, and other entities, supporting both candidate discovery and explanation [7]. Separate studies have explored emotion from facial images and attention over review text [8, 9, 10]. These techniques address different gaps, but also introduce distinct data requirements, error modes, and privacy costs.

This paper proposes a confidence-aware cascade that combines complementary signals while preserving a fallback when any component is unavailable. Its contribution is a system design and evaluation protocol, not a claim of measured improvement. The design treats ratings and viewing actions as primary behavioral evidence; metadata and graph relations as item evidence; and text or transient mood as optional contextual evidence. The remainder reviews the supplied literature, defines the framework, describes its ranking and privacy rules, and presents an evaluation plan.

2 Related Work

2.1 Hybrid filtering and interaction data

The supplied studies describe demographic, content-based, user-based, and item-based filtering as complementary approaches. A machine-learning implementation by Rao et al. combines demographic filtering, content similarity, and collaborative methods, including item-based filtering and singular value decomposition [1]. Malik likewise motivates a hybrid of content-based and collaborative filtering, using movie attributes and historical interactions [2]. Agrawal presents a hybrid that combines content and collaborative filtering with a support vector machine and genetic algorithm, reporting comparative evaluation across three datasets [3]. These studies support the design choice to separate candidate generation from ranking so different evidence can contribute without forcing a single algorithm to serve every case.

Pu and Hu report an 81% accuracy for their hybrid recommender and state that it exceeds their traditional content-based, item-based, and user-based baselines [4]. Because the supplied paper's dataset, split, metric definition, and baseline settings are specific to that experiment, the reported percentage should not be generalized to the framework proposed here or compared directly with metrics from other papers. Jadhav et al. describe recommendations based on user preference and similarity to other users and include top-rated listings [5]. Together, the work motivates hybridization and explicit cold-start handling, while leaving open how confidence should govern the contribution of each signal.

2.2 Sentiment and emotion as contextual signals

Sentiment can add preference information that is not captured by a star rating. Chu et al. propose an attention-based recommendation algorithm that extracts emotion from review text and uses a score-correction gate before collaborative filtering [6]. The cascading hybrid model by Raghavendra et al. uses movie tweets and combines sentiment-aware information with hybrid recommendation [7]. Facial emotion has also been explored as an input: Chauhan et al. discuss facial-expression recognition for movie recommendations [8], and Zhang reports a ResNet-38 emotion classifier with 64.02% test accuracy on FER-2013 in a movie and music recommendation setting [9]. These results suggest possible context features, but facial-expression classification accuracy is not equivalent to recommendation quality. Context should therefore be optional, permission-based, and evaluated separately from ranking effectiveness.

2.3 Knowledge graphs and scalable recommendation

Luo et al. design a movie recommendation system based on a knowledge graph, emphasizing connected movie attributes and graph visualization [10]. Graph relations can supply paths such as movie-to-director-to-other-movie and movie-to-genre-to-similar-title, which can support candidate generation and human-readable explanations. Zhou et al. describe item-based collaborative filtering implemented with MapReduce to handle large data sets [11]. This motivates an architecture in which offline computation

builds embeddings, neighborhoods, and graph indexes, while online serving performs bounded candidate retrieval and ranking.

2.4 Synthesis and design requirements

Across the supplied papers, the recurring limitation is dependence on a narrow evidence source: interaction methods need sufficient histories; content methods need informative metadata; sentiment methods need text; and facial emotion methods require an image signal that is particularly sensitive. The synthesis yields four requirements: (1) support multiple candidate sources; (2) attenuate uncertain or missing evidence rather than treating it as a preference; (3) provide a cold-start path; and (4) make context collection controllable. These requirements define the framework below.

3 Proposed Framework

3.1 Architecture and data flow

The architecture has an offline preparation path and an online recommendation path. Offline preparation validates catalog metadata, maps entities into a movie knowledge graph, computes content representations, and derives collaborative neighborhoods or latent factors from historical interactions. Online serving reads a user profile and current session only when available, retrieves candidates from the content, graph, and collaborative channels, deduplicates them, applies eligibility filters, and ranks the remaining items. Contextual sentiment or mood may adjust the ranking only when the user has enabled that feature and the signal passes a reliability threshold. The design is illustrated in Figure 1.

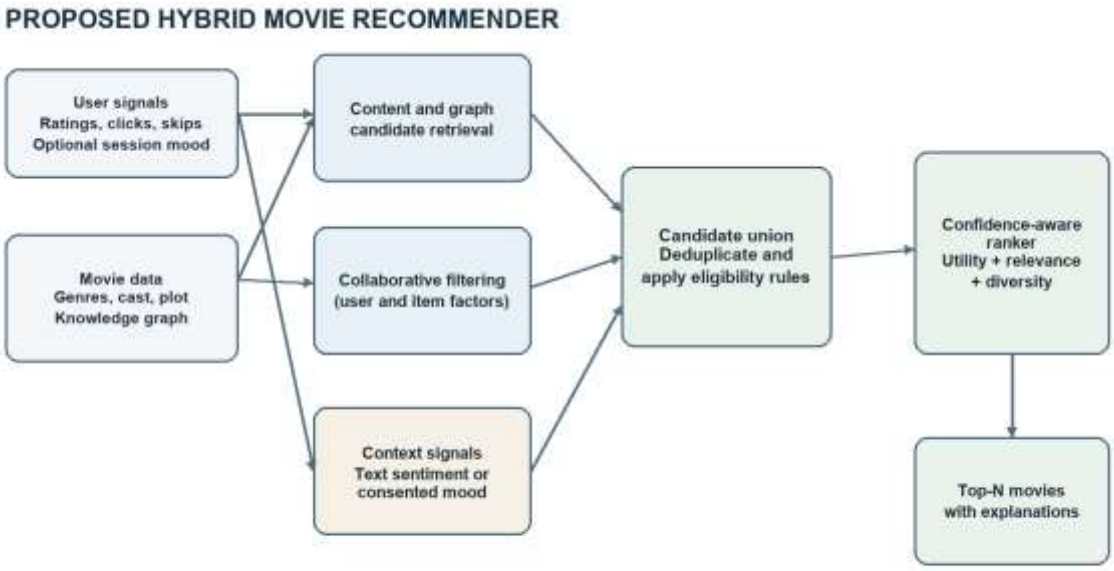


Figure 1. Proposed Confidence-Aware Hybrid Recommendation Architecture

3.2 Candidate generation

Let U denote users and M denote movies. The interaction matrix R contains explicit ratings where available and weighted implicit events such as a completed watch, a short preview, a skip, or a save. Event weights must be calibrated on validation data and should distinguish exposure from preference: a title that was never shown is not a negative example. Three candidate sets are retrieved: C_content from metadata similarity to the user's positive profile; C_graph from short, typed paths in the knowledge graph; and

C_CF from item neighborhoods or latent user-item factors. A popularity-based set C_pop provides a fallback for a new user or a channel with no usable evidence. The union is bounded to control serving cost.

The content representation can combine genre, cast, director, language, release period, and plot features. Categorical fields are normalized and multi-valued fields are treated as sets. Text fields may be represented with term weighting or learned embeddings, but training and serving must use the same versioned feature pipeline. Graph retrieval uses allowed relation types and bounded path length so that candidates remain explainable and graph expansion does not dominate latency.

3.3 Confidence-aware ranking

For a user u and candidate movie m, each channel returns a normalized relevance score s_k(u,m), bounded between zero and one, and a confidence c_k(u,m), also bounded between zero and one, where k identifies content, graph, collaborative, or context evidence. Confidence reflects evidence quantity and quality, such as interaction count, metadata completeness, graph-path support, or classifier calibration. The proposed score is a normalized weighted sum:

$$\text{Score}(u,m) = [\Sigma_k\, w_k\, c_k(u,m)\, s_k(u,m)] \,/\, [\Sigma_k\, w_k\, c_k(u,m) + \varepsilon] + \lambda\, D(u,m) \quad (1)$$

Here w_k are validation-tuned channel weights, ε prevents division by zero, and D is a diversity adjustment that penalizes redundancy among the already selected top-ranked films. The context weight is set to zero when consent is absent, the signal is unavailable, or classifier confidence is below its calibrated threshold. If all personalized channel confidences are near zero, the system falls back to a quality-aware popularity list filtered by availability and basic user settings. The formula is a design specification; its weights and thresholds require empirical selection.

A practical cascade first applies eligibility constraints, then retrieves a broad candidate pool, then computes the more expensive collaborative and graph features, and finally applies ranking and diversity. This ordering allows cost limits to be enforced without changing the meaning of a recommendation. A list-level reranker can balance relevance and diversity after item scores have been computed.

3.4 Feedback and explanations

After serving a list, the system records only the events needed for recommendation and evaluation under the platform's retention policy. Explicit ratings and deliberate saves provide stronger evidence than impressions; a skip is weak evidence unless exposure and viewing context are known. An explanation can cite the signal that materially contributed to the score, for example "recommended because you liked films by this director" or "viewers with similar ratings also liked this title." Context-derived explanations should not reveal sensitive inferred emotion. Figure 2 presents the request and feedback loop.

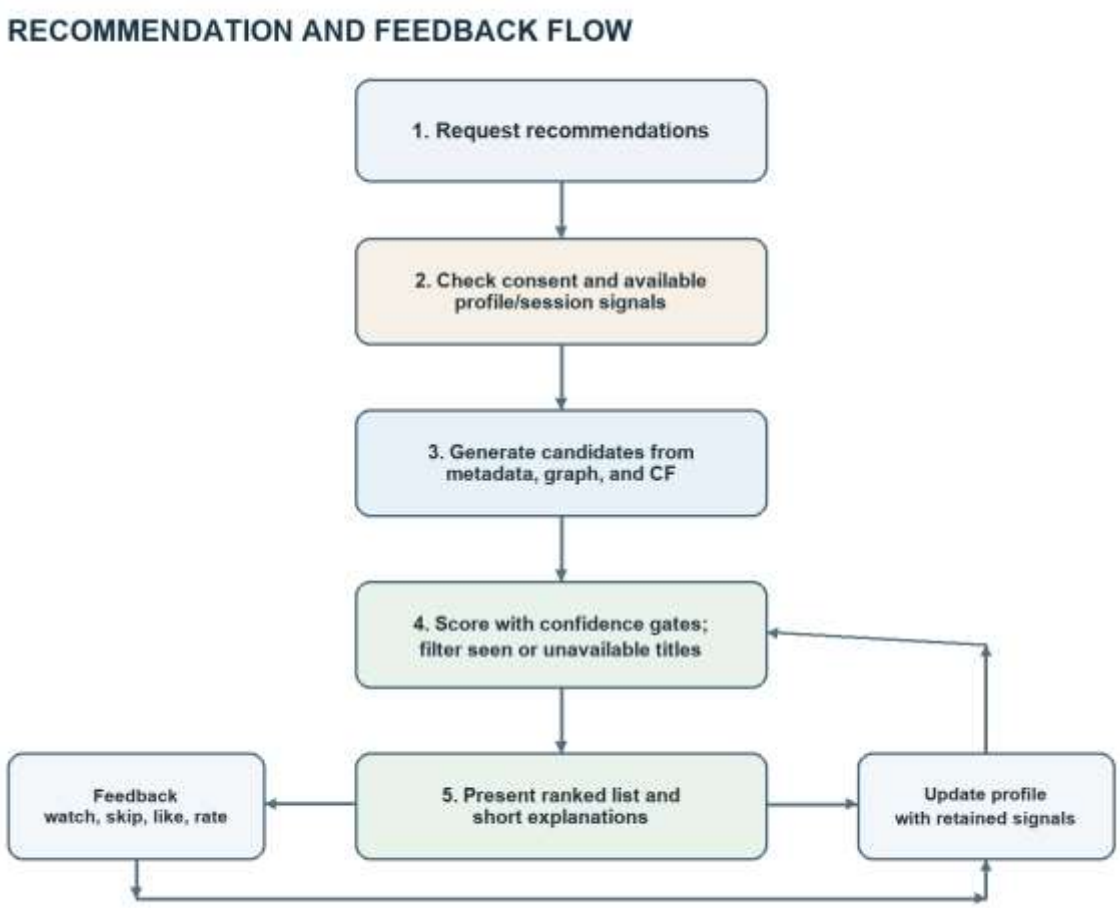


Figure 2. Request, Ranking, and Feedback Flow

4 Privacy and Reliability Considerations

The system should collect the least sensitive signal that can meet the task. Ratings, saves, and watch events are sufficient for the default profile. Review text is processed only when the user has chosen to provide it, and raw text need not be retained after feature extraction unless a separate purpose and retention rule justify it. Facial emotion is not required by the proposed core architecture. If offered as an optional session input, it should be processed transiently, require explicit opt-in, provide a clear off switch, and avoid storing face images or inferred emotion as a persistent profile by default. The user must be able to request recommendations without camera access.

Reliability controls include calibration of sentiment and emotion probabilities, monitoring for demographic and language-specific error, and a neutral fallback when context is uncertain. Feedback loops can amplify exposure bias: items shown more often receive more interactions and may appear increasingly relevant. Evaluation should therefore distinguish ranking quality from popularity effects, report coverage and novelty, and use exposure-aware logging. Knowledge graph relations and item metadata should also be audited for stale, missing, or misleading links.

5 Evaluation Plan

The proposal should be evaluated against content-based, item-based collaborative, user-based collaborative, graph-only, and non-contextual hybrid baselines. A public movie-rating collection such as

MovieLens can supply explicit interactions, while catalog metadata can supply genres, cast, and descriptions. If a text-sentiment component is evaluated, it requires a documented review corpus with appropriate usage rights; facial emotion should be a separate opt-in study and must not be silently joined to historical ratings. Train, validation, and test splits should be temporal to reduce leakage from future interactions. Hyperparameters and confidence thresholds should be selected on validation data only.

Ranking quality should be reported with precision@K, recall@K, normalized discounted cumulative gain, and mean average precision. Rating prediction may be reported separately with root mean squared error or mean absolute error when explicit ratings are the target. Coverage, catalog coverage, diversity, novelty, and average serving latency should be reported alongside accuracy. Cold-start performance should be measured in separate cohorts for new users and new movies. Confidence intervals should be estimated across users or repeated temporal folds, and ablations should remove the graph, collaborative, sentiment, and diversity components one at a time.

A fair comparison requires a common preprocessing pipeline, identical candidate availability, consistent filtering rules, and the same split for all baselines. Results should not combine the 81% accuracy reported by Pu and Hu [4] or the FER-2013 emotion-classification accuracy reported by Zhang [9] with recommendation metrics from a different task. No such experiment was conducted for this paper; the plan specifies what evidence would be needed to assess the proposed design.

6 Discussion

The framework's main design advantage is graceful degradation. For a new user, content preferences collected through optional onboarding and a popularity fallback can provide an initial list. For a new movie, metadata and graph connections can retrieve candidates before interactions accumulate. When sentiment or mood is missing, the recommender continues with behavioral, content, and graph evidence. The confidence gate prevents a weak signal from having the same influence as a well-supported one. The design also makes the trade-off between richer context and privacy visible instead of treating every available feature as mandatory.

The proposal has limitations. Confidence estimates can themselves be miscalibrated, and a weighted score does not guarantee fairness or user satisfaction. A knowledge graph is costly to curate and may encode incomplete or culturally narrow associations. Text sentiment can confuse plot description, sarcasm, and the reviewer's stance toward a film. Facial expression does not reliably identify a person's preference or mood. Lastly, offline ranking metrics cannot establish whether users feel helped or whether recommendations improve long-term satisfaction. These points require user studies and ongoing monitoring.

7 Conclusion

A modular hybrid recommender has been specified that combines metadata, collaborative filtering, knowledge-graph relations, and optional context through a confidence-aware cascade. It provides a cold-start fallback, explains recommendations using non-sensitive evidence, and limits context to signals that are both reliable and permitted. The supplied literature supports the relevance of each component but does not validate this particular integration. A temporal, leakage-controlled evaluation with ranking, coverage, diversity, latency, and privacy measures is therefore the next step before any performance claim is made.

References

1. Narendra Kumar Rao, Nagendra Panini Challa, Ranjana R, S. Sreenivasa Chakravarthi, "Movie Recommendation System using Machine Learning," Proceedings of the 4th International Conference on Inventive Research in Computing Applications, IEEE, 2022, pp. 710-716, DOI: 10.1109/ICIRCA54612.2022.9985512.
2. Sonika Malik, "Movie Recommender System Using Machine Learning," EAI Endorsed Transactions on Creative Technologies, vol. 9, no. 3, 2022, DOI: 10.4108/eetct.v9i3.2712.
3. Shreya Agrawal, "An Improved Approach for Movie Recommendation System," Proceedings of the International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), IEEE, 2017, pp. 336-342.
4. Qingna Pu, Bin Hu, "Intelligent Movie Recommendation System Based on Hybrid Recommendation Algorithms," Proceedings of the 2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics, IEEE, 2023, DOI: 10.1109/AIKIIE60097.2023.10389982.
5. Rupali Jadhav, Snehal Inamdar, Sakshi Karanjkar, Ritesh Karankal, Riya Kate, "Implementing a Movie Recommendation System in Machine Learning Using Hybrid Approach," International Journal for Research in Applied Science & Engineering Technology, vol. 11, issue 5, May 2023, DOI: 10.22214/ijraset.2023.53204.
6. Yan Chu, Shuhao Qi, Yue Yang, Chenqi Shan, Lina Wang, Zhengkui Wang, "An Attention-based Recommendation Algorithm," Proceedings of the 2019 IEEE International Conference on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking, IEEE, 2019, DOI: 10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00218.
7. Raghavendra C. K., Mahantesha U., Prashanth J., "Cascading Based Hybrid Recommendation Model," Proceedings of the 5th International Conference on Mobile Computing and Sustainable Informatics, IEEE, 2024, DOI: 10.1109/ICMCSI61536.2024.00033.
8. Shavak Chauhan, Rajdeep Mangrola, D. Viji, "Analysis of Intelligent Movie Recommender System from Facial Expression," Proceedings of the 5th International Conference on Computing Methodologies and Communication, IEEE, 2021, pp. 1454-1461, DOI: 10.1109/ICCMC51019.2021.9418421.
9. Jingye Zhang, "Movies and Pop Songs Recommendation System by Emotion Detection through Facial Recognition," Journal of Physics: Conference Series, vol. 1650, 032076, 2020, DOI: 10.1088/1742-6596/1650/3/032076.
10. Lixia Luo, Zuoliu Huang, Qitao Tang, "Research and Implementation of Movie Recommendation System Based on Knowledge Graph," Proceedings of the 4th International Conference on Computer, Big Data and Artificial Intelligence, IEEE, 2023, DOI: 10.1109/ICCBD-AI62252.2023.00109.
11. Tianqi Zhou, Lina Chen, Jian Shen, "Movie Recommendation System Employing the User-based CF in Cloud Computing," Proceedings of the 2017 IEEE International Conference on Computational Science and Engineering and IEEE International Conference on Embedded and Ubiquitous Computing, IEEE, 2017, DOI: 10.1109/CSE-EUC.2017.194.