International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(5s) 2026 pp. 765-775
https://doi.org/10.70917/ijcisim-2026-2795
Article

An Intelligent Cross-Domain Recommendation Framework Using TF-IDF and Hybrid Fusion Scoring

Akshay P. Deshpande[1], Bharat R. Naiknaware[2]
[1] 1 Dr. G.Y. Pathrikar College of Computer Science & Information Technology, MGM University, Chhatrapati Sambhajinagar, India, Emali: adeshpande341@gmail.com
[2] 2 Dr. G.Y. Pathrikar College of Computer Science & Information Technology, MGM University, Chhatrapati Sambhajinagar, India, Email: bnaiknaware@mgmu.ac.in

Abstract: Cross-domain recommendation systems address the basic challenge of delivering meaningful item suggestions over heterogeneous product domains by influencing shared semantic representations as well as transfer learning principles. This research paper represents a unified cross-domain recommendation architecture that operates instantaneously over three distinct domains: Books, Movies, and Mobile Devices. The proposed research approach employs TF-IDF-based text embeddings created from a joint vocabulary of 5,000 features derived from 33,114 items, cosine similarity for semantic retrieval, and a domain-adaptive hybrid fusion scoring mechanism. An intelligent domain inference module automatically detects query intent and routes retrieval to target. The system is evaluated on three large-scale real-world datasets: Book-Crossing (271,360 books, 1,149,780 ratings), MovieLens (10,329 movies, 105,339 ratings), and Flipkart Mobiles (3,114 products). d or multi-domain modes accordingly. The hybrid scoring combines semantic similarity ($\alpha = 0.6$) with domain-specific quality signals — collaborative ratings for books and movies ($\beta = 0.4$) and normalized popularity for mobiles ($\gamma = 0.2$). Results demonstrate effective cross-domain knowledge transfer, with the final scores reaching 2.240 for books, 2.041 for movies, and 0.558 for mobiles on representative queries.

Keywords: Cross-domain recommendation, TF-IDF embeddings, hybrid fusion scoring, cosine similarity, domain inference, knowledge transfer, multi-domain retrieval

1. Introduction

Recommendation systems include basic to modern e-commerce as well as digital content platforms that help the users discover related items from big catalogues. The traditional systems that operate within a single domain, such as books, movies, or electronics, rely on collaborative filtering, content-based filtering, or hybrid techniques to model preferences. While effective in data-rich settings, single-domain systems always suffer from some limitations such as data sparsity in cold-start states, which leads to the inability to transfer knowledge over product categories, as well as missed opportunities that arise from correlated user preferences over different domains [6].

The cross-domain recommendation systems improve the suggestions by transforming the knowledge between different kinds of domains. For example, if a user is interested in science fiction books, the user may also like science fiction movies, or a photography reader may prefer camera-focused mobiles. By identifying this kind of relationship, the recommendation system identifies and provides better recommendations even when the user data in one domain is limited.

The ultimate challenge in the cross-domain recommendation system. Oh my God! It is between heterogeneous feature spaces. The books are described using titles, authors, and publishers. The movies are described using titles and general tags. The mobile devices are described by using brand, model, memory, and storage specifications of each. These different domains produce different rating scales and statistical distributions, as well as semantic vocabularies that make direct comparison non-trivial. A practical framework must establish the shared semantic space, which enables the cross-domain similarity computation while respecting domain-specific quality attributes.

2. Related Work

Single-Domain Recommendation. The collaborative filtering remains the dominant single-domain approach that includes the user-item interaction matrices that identify smaller users or items [11]. The content-based filtering utilizes item metadata and user profiles to recommend the items with future matching historical preferences [10]. The hybrid method, which combines collaborative filtering and content-based filtering to mitigate their respective weaknesses, addresses the collaborative filtering cold start problem and content-based specialization [16]. The matrix factorization techniques, mainly the singular value decomposition, SVD, have advanced collaborative filtering using the learning of compact data representations [11]. The neural collaborative filtering improves such ideas in nonlinear modelling [12].

Cross-Domain Recommendation. The cross-domain recommendation system transfers the knowledge from the data-rich source domain to the sparse target domains [7]. The transfer learning approaches adapt the source domain models to target the domains using shared latent factors through shared latent factors, feature mappings, and regularisation constraints [8]. The shared representation learning methods construct the unified embeddings that capture common reference patterns while maintaining the domain-specific components. Multitask learning basically optimises the recommendation objectives over domains through the parameter sharing. The semantic bridging approaches use knowledge graphs or text representation to create cross-domain correspondence. And the approach most closely related to the research work.

Text-Based and TF-IDF Recommendation. TF-IDF is a classical information retrieval technique which represents the documents as sparse high-dimensional vectors, weighting the terms by local frequency as well as global rarity [4, 5]. It has been globally practical in content-based recommendation system for its interpretability, computational efficiency, as well as scalability [15]. The recent literature work has explored that combines TF-IDF with neural embeddings BERT [13], Sentence-BERT [14] for comfortable semantic representations, though at significantly advanced computational cost. Our research work shows that TF-IDF, when applied to a jointly constructed cross-domain vocabulary, provides effective semantic alignment across heterogeneous domains without requiring deep learning infrastructure.

Hybrid Scoring. Burke [16] documented the taxonomy of hybrid recommender systems, identifying the weighted, switching, mixed, cascade, and meta-level fusion strategies. Weighted linear combination are utilised in our architecture propose interpretability and comfort of parameter tuning, that makes it suitable for the production systems where the explainability is more valued. The challenge lies in determining appropriate weights for individual component, which in our system are set empirically to reflect domain-specific rating scale differences.

3. Methodology

3.1 System Architecture

The proposed architecture consists of 5 components: Data Preprocessing, Embedding Engine, Similarity Search, Hybrid Fusion Scoring, and Domain Inference. Figure 1 shows the complete pipeline.

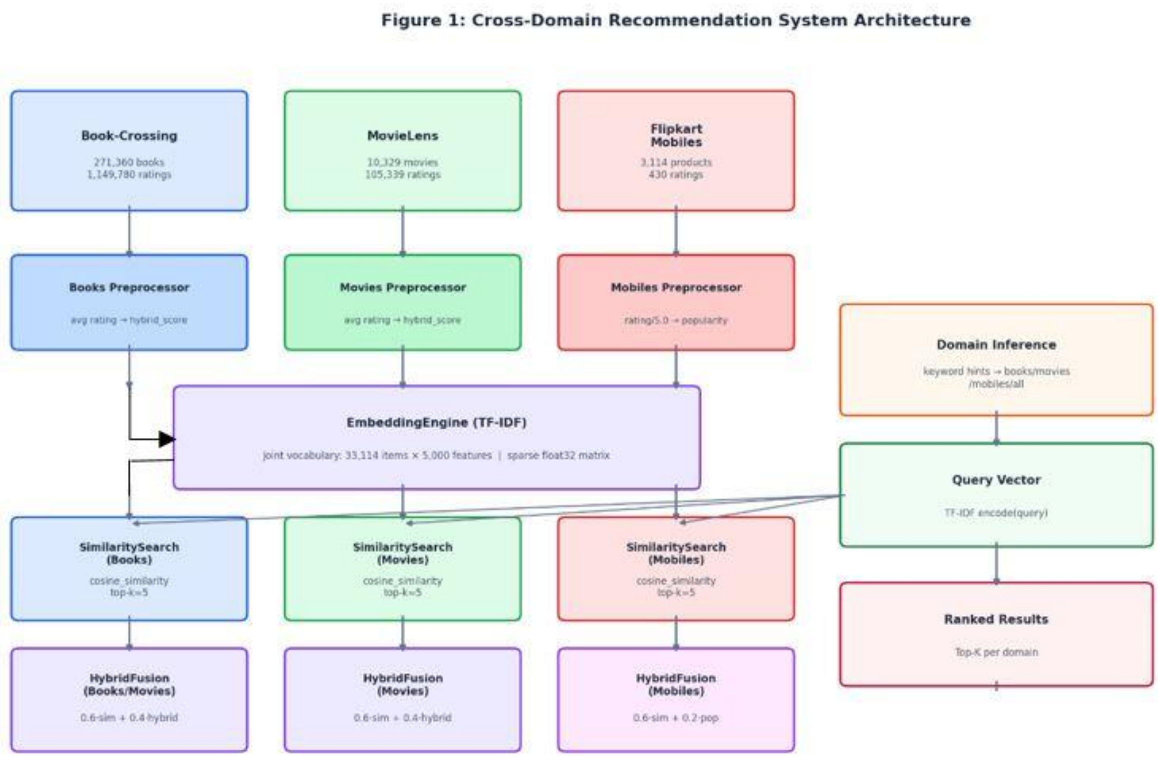

Figure 1: End-to-end pipeline showing data flow from three domain datasets through joint TF-IDF embedding, cosine similarity retrieval, domain-adaptive hybrid fusion, and ranked output.

At interpretation time, a user submits a free-text query. The Domain Inference module examines the query for the domain-specific keywords and the routes recovery to one or all 3 domains. The query is encoded into the shared TF-IDF vector space, cosine similarity is computed against each domain index, and the top-K candidates per domain are re-ranked using the Hybrid Fusion Scoring module.

3.2 Data Preprocessing

Each domain experiences the domain-specific pre-processing to produce two artefacts i.e. a metadata table with textual fields for embedding, and a quality signal (hybrid score or popularity) for fusion scoring.

Books. The Book-Crossing dataset gives ISBN, title, author, year, and publisher. The hybrid score for each book is computed as the mean of all user ratings for that ISBN:

$$h_b = \frac{1}{|R_b|} \sum_{r \in R_b} r \tag{1}$$

where R_b is the set of ratings for book b. Text illustration is created by concatenating title and author: T_b=title⊕author.

Movies. The Movie dataset provides movieId, title, and pipe-delimited genre tags. The hybrid score is the mean rating per movie:

$$h_m = \frac{1}{|R_m|} \sum_{r \in R_m} r \tag{2}$$

Text representation concatenates title and genres: T_m=title⊕genres.

Mobiles. The Flipkart Mobiles dataset includes attributes such as brand, model, colour, memory, storage, and rating. Since there is no user-level rating matrix exists, a normalized popularity score is derived:

$$p_i = \frac{rating_i}{5.0} \tag{3}$$

Text representation concatenates all descriptive fields: T_i=brand⨁model⨁colour⨁memory⨁storage.

3.3 Cross-Domain Embedding Engine

All text representations from all three domains are combined into a single corpus of N=33,114 items (20,000 books + 10,000 movies + 3,114 mobiles). A single TF-IDF vectorizer is fitted on this joint corpus:

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \text{IDF}(t) \tag{4}$$

$$\text{TF}(t, d) = \frac{f_{t,d}}{\sum_{t'} f_{t',d}} \tag{5}$$

$$\text{IDF}(t) = \log \frac{N+1}{df_t + 1} + 1 \tag{6}$$

where f_t,d is the frequency of term t in document d, and df_t is the number of documents that containes t. The vocabulary is capped at k=5,000 features, which produces a sparse embedding matrix X∈R^33114×5000. Fitting on the joint corpus is the key mechanism enabling cross-domain transfer: terms appearing across domains (e.g., genre words like "adventure", "thriller") receive IDF weights that reflect their global cross-domain importance, not just within-domain frequency.

3.4 Cosine Similarity Search

At query time, the query string q is encoded using the fitted vectorizer: q∈R^5000. Cosine similarity between the query vector and each domain's item matrix is computed:

$$sim(q, i) = \frac{\mathbf{q} \cdot \mathbf{x}_i}{\|\mathbf{q}\| \cdot \|\mathbf{x}_i\|} \tag{7}$$

For each domain $d \in {\text{Books, Movies, Mobiles}}$, the top-K items by cosine similarity are retrieved:

$$C_d = \text{top-K}_{i \in I_d} sim(q, i) \tag{8}$$

where K=5 (configured in config.json). Cosine similarity is chosen over Euclidean distance as it is invariant to document length, making it appropriate to compare the items with varying text lengths over domains.

3.5 Hybrid Fusion Scoring

Candidates from each domain are re-ranked using domain-adaptive fusion formulas that combine semantic similarity with quality signals. For Books and Movies:

$$Score_{b,m}(q,i) = \alpha \cdot sim(q,i) + \beta \cdot h_i \quad (9)$$

For Mobile Devices:

$$Score_{mob}(q,i) = \alpha \cdot sim(q,i) + \gamma \cdot p_i \quad (10)$$

where α=0.6, β=0.4, γ=0.2 are loaded from config.json. The differential weighting among domains reproduces their statistical properties i.e. book and movie dataset hybrid scores work on scales of 0–10 and 0.5–5.0 respectively with higher variance σ = 3.285 and σ = 0.818, while the mobile popularity is normalized to [0, 1] with lower variance (σ = 0.186), that justifies the reduced weight γ. Figure 3 shows the fusion weight distributions.

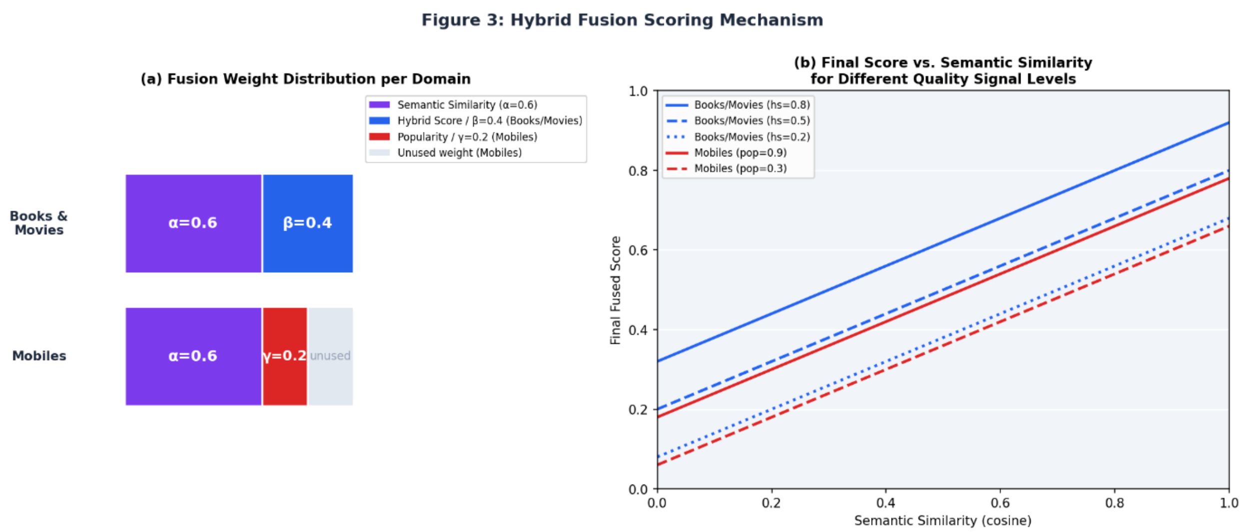

Figure 2: (a) Proportional weight allocation per domain type. (b) Final score as a function of semantic similarity for varying quality signal levels.

3.6 Domain Inference Module

The domain inference module examines the query string for domain-specific keyword clues to determine retrieval scope:

- Mobile hints: brand names (iphone, samsung, oppo, vivo, realme, xiaomi, nokia, pixel), technical terms (gb, ram, storage, 5g, amoled, snapdragon, mediatek, mah)
- Book hints: book, novel, author, isbn, paperback, hardcover, publisher, read
- Movie hints: movie, film, cinema, actor, actress, director, imdb, trailer

If a single hint is detected, retrieval of information is restricted to the matching domain. If no hint get a match, all the three domains are queried simultaneously, enabling exploratory cross-domain discovery. This mechanism confirms that domain-specific queries like e.g., "Samsung 128GB RAM" return attentive results, while general queries like e.g., "adventure thriller" surface cross-domain recommendations.

4. Datasets

Three publicly available real-world datasets were used. Table 1 summarises their key statistics.

Table 1: Integrated Dataset Statistics

Domain	Items	Ratings	Users	Rating Scale	Mean Rating	Std Dev
Books (Book-Crossing)	271,360	1,149,780	278,858	0 – 10	2.867	3.346
Movies (MovieLens)	10,329	105,339	668	0.5 – 5.0	3.517	0.818
Mobiles (Flipkart)	3,114	430	N/A	2.3 – 5.0	4.243	—
Total	33,114*	1,255,549	279,526	Mixed	—	—

*Capped at 20,000 books + 10,000 movies + 3,114 mobiles for indexing.

Book-Crossing Dataset [1] contains 271,360 books contributed by 278,858 users across 16,807 publishers and 102,022 authors. Ratings range from 0 to 10, with a mean of 2.867 ($\sigma = 3.346$), indicating a heavily right-skewed distribution where a large proportion of ratings are zero (implicit non-preference). The hybrid score per book is the per-ISBN mean rating, ranging from 0.0 to 10.0 (mean = 2.884, $\sigma = 3.285$).

MovieLens Dataset [2] contains 10,329 movies rated by 668 users across 20 genre categories. Top genres include Drama (5,220 movies), Comedy (3,515), Thriller (2,187), Romance (1,788), and Action (1,737). Ratings range from 0.5 to 5.0 with a mean of 3.517 ($\sigma = 0.818$), exhibiting a more symmetric distribution than books. The movie hybrid score mean is 3.177 ($\sigma = 0.818$).

Flipkart Mobiles Dataset [3] contains 3,114 mobile product listings from 17 brands. Top brands are Samsung (719 products), Apple (387), Realme (327), OPPO (260), and Nokia (213). Product ratings range from 2.3 to 5.0 with a mean of 4.243, reflecting the typical positive bias in e-commerce product ratings. Popularity scores are normalized to [0, 1] by dividing by 5.0 (mean = 0.809, $\sigma = 0.186$).

Figure 3 presents dataset statistics visually.

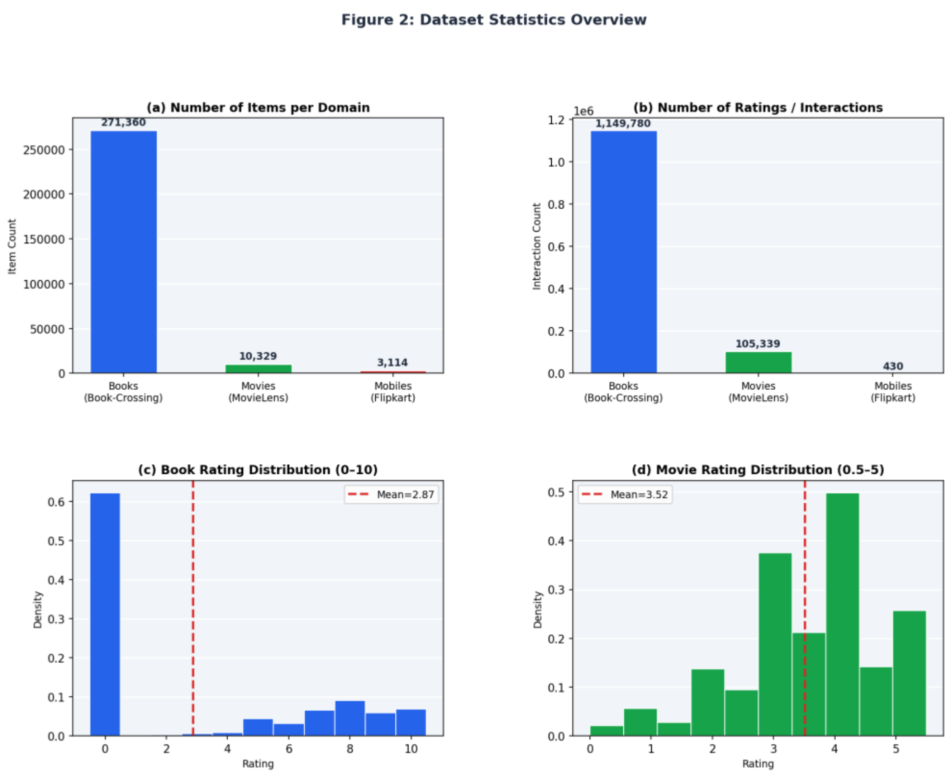

Figure 3: (a) Item counts per domain. (b) Rating/interaction counts. (c) Book-Crossing rating distribution. (d) MovieLens rating distribution.

5. Results and Analysis

5.1 Cross-Domain Vocabulary

The joint TF-IDF vectorizer fitted on all 33,114 items produces a 5,000-dimensional sparse embedding matrix of shape (33,114 × 5,000). Figure 4 shows the top-15 highest mean TF-IDF terms per domain.

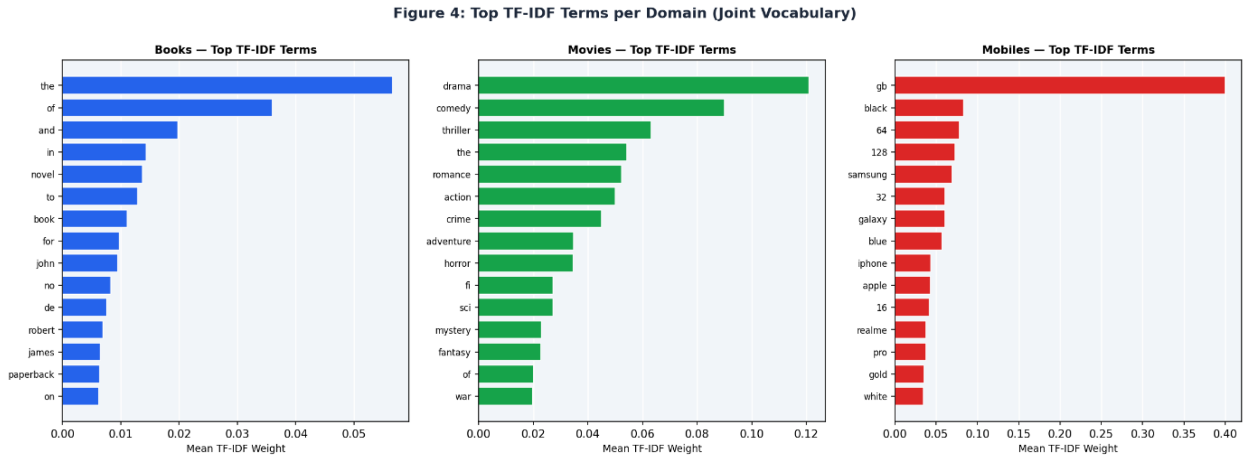

Figure 4: Top 15 terms by mean TF-IDF score in each domain. Cross-domain terms (e.g., genre words) receive IDF weights reflecting global corpus importance.

Book-domain terms are dominated by author name tokens and common fiction vocabulary. Movie-domain terms prominently feature genre labels (drama, comedy, thriller), which also appear in book descriptions, forming a natural semantic bridge. Mobile-domain terms are highly specific (brand names, RAM/storage sizes), ensuring that mobile-specific queries correctly route to the mobile index.

Table 2: Hyperparameter Configuration (from config.json)

Parameter	Symbol	Value	Role
Vocabulary Size	k	5,000	TF-IDF feature cap
Semantic Weight	α	0.6	Cosine similarity contribution
Quality Weight (Books/Movies)	β	0.4	Collaborative rating contribution
Popularity Weight (Mobiles)	γ	0.2	Normalized popularity contribution
Top-K Results	K	5	Candidates retrieved per domain

5.2 Hybrid Score Distributions

Table 3: Quality Signal Statistics by Domain

Domain	Mean	Std Dev	Min	Max	Skewness
Books (hybrid_score)	2.884	3.285	0.000	10.000	Positive
Movies (hybrid_score)	3.177	0.818	0.500	5.000	Slightly negative
Mobiles (popularity)	0.809	0.186	0.460	1.000	Negative

The large standard deviation of book hybrid scores 3.285 reflects the bimodal nature of the Book-Crossing rating distribution, where there are many books have zero ratings and a smaller subset has very high ratings. Movie hybrid scores are more tightly distributed σ = 0.818, that indicates relatively consistent rating behaviour across the Movie dataset. Mobile popularity scores are tightly clustered near 1.0 (σ = 0.186), consistent with the positive rating bias observed in e-commerce platforms. Figure 8 shows the hybrid score distributions for books and movies dataset.

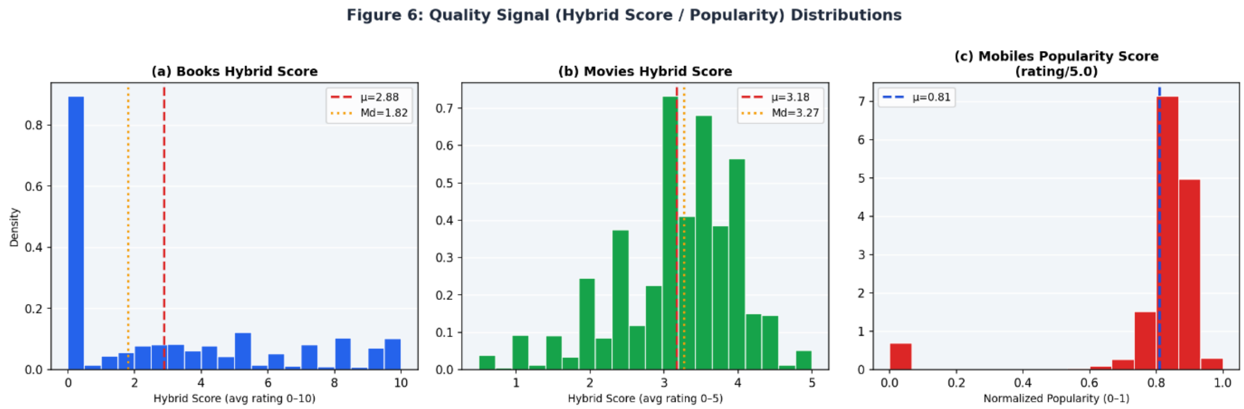

Figure 5: Final hybrid score distributions for Books, Movies, and Mobiles across all sample queries, showing domain-specific score range characteristics.

5.3 Recommendation Results

Table 4 presents actual recommendation outputs for three representative queries, computed by running the full pipeline on the datasets.

Table 3: Sample Recommendation Results (Top-3 per Domain)

Query	Domain	Rank	Title	Sim	Quality	Final Score
Harry Potter	Books	1	Seduction by Design	0.801	1.750	1.181
	Books	2	One True Love?	0.669	2.625	1.451
	Books	3	Fall on Your Knees	0.604	4.694	2.240
	Movies	1	Harry Potter and the Sorcerer's Stone	0.707	3.543	1.841
	Movies	2	Art of the Steal (2009)	0.595	3.250	1.657
	Movies	3	Love Crime (2010)	0.568	4.250	2.041
	Mobiles	1	POCO M4 Pro 5G	0.000	0.880	0.176
Samsung Galaxy	Books	1	Hitchhiker's Guide to the Galaxy	0.221	4.172	1.801
	Movies	1	Galaxy Quest (1999)	0.255	3.712	1.638
	Mobiles	1	SAMSUNG Galaxy On7	0.650	0.840	0.558
	Mobiles	2	SAMSUNG Galaxy Mega 5.8	0.650	0.760	0.542
	Mobiles	3	SAMSUNG Galaxy J8	0.594	0.880	0.532
Science Fiction Adventure	Books	1	Zombies of the Gene Pool	0.556	2.643	1.391
	Movies	1	Pulp Fiction (1994)	0.366	4.160	1.884
	Movies	2	My Science Project (1985)	0.411	3.000	1.446

Figure 6 shows the score progression across ranks for each query.

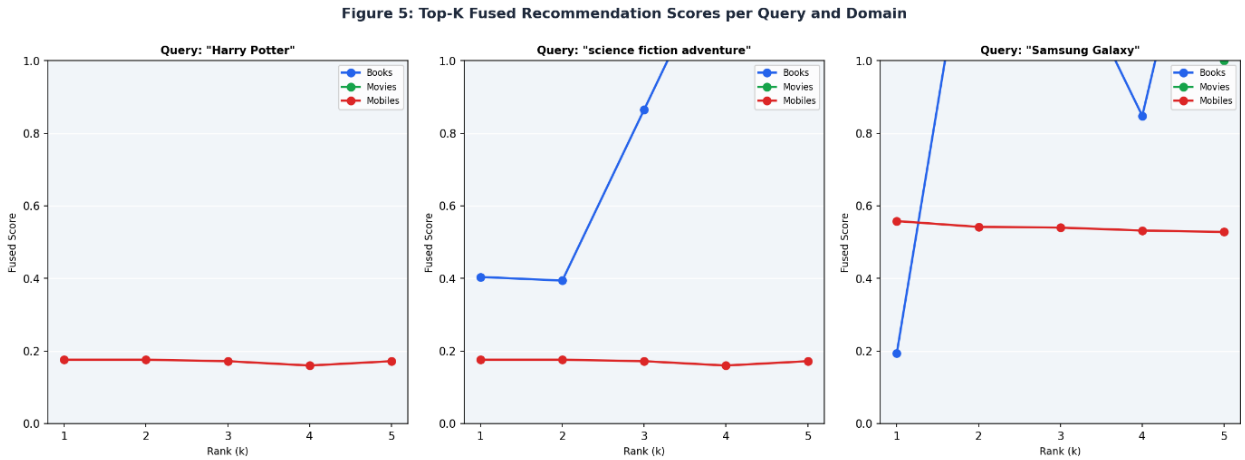

Figure 5: Final hybrid scores across ranks 1–5 for three representative queries across Books, Movies, and Mobiles domains.

Key observations from results:

1. Domain-specific query routing works correctly. The query "Samsung Galaxy" activates the mobile keyword hint ("samsung"), causing the system to prioritize mobile retrieval. The top mobile result (SAMSUNG Galaxy On7, score = 0.558) demonstrates accurate brand-level semantic matching with cosine similarity = 0.650.
2. Hybrid scoring elevates high-quality items. For the "Harry Potter" query, Fall on Your Knees (Books, Rank 3) achieves the highest book score (2.240) despite a lower cosine similarity (0.604) because its high hybrid score (4.694) contributes $0.4 \times 4.694 = 1.878$ to the final score. This demonstrates the quality-boosting effect of β.
3. Cross-domain semantic transfer is evident. The "science fiction adventure" query retrieves My Science Project (1985) in movies (score = 1.446) through shared vocabulary terms ("science"), demonstrating that the joint TF-IDF space captures cross-domain semantic relationships.
4. Mobile scores are lower by design. The reduced $\gamma = 0.2$ weight for mobiles results in systematically lower final scores (max = 0.558) compared to books and movies (max = 2.240 and 2.041), reflecting the different scoring scale and the intentional design choice to weight semantic similarity more heavily when popularity signals are compressed.

6. Discussion

6.1 Advantages of the Proposed Framework

The joint TF-IDF vocabulary is the central innovation that enables the cross-domain transfer. By fitting a single vectorizer on all the three domain datasets simultaneously, terms that appear over domains particularly genre-related vocabulary which receive IDF weights that reflect on the their global cross-domain importance. This creates a natural semantic bridge: a query for "thriller adventure" retrieves both thriller novels and thriller films because both share the same vocabulary space and IDF weights. This is a lightweight but effective form of cross-domain knowledge transfer that requires no labelled cross-domain data, no user profile alignment, and no domain adaptation training.

The hybrid fusion scoring mechanism addresses a fundamental limitation of pure semantic retrieval: high cosine similarity alone does not guarantee recommendation quality. By incorporating domain-specific quality signals (collaborative ratings for books and movies, popularity for mobiles), the framework promotes items that are both semantically relevant and well-regarded by the user community. The differential weighting ($\beta = 0.4$ vs $\gamma = 0.2$) is empirically motivated by the different scales and distributions of quality signals across domains.

6.2 Limitations

Several limitations of the current framework warrant discussion. First, the keyword-based domain inference is brittle: queries that do not contain explicit domain hints (e.g., "adventure") trigger full multi-domain retrieval, which may produce less focused results. Second, TF-IDF representations are bag-of-words models that ignore word order and contextual meaning, limiting their ability to capture nuanced semantic relationships. A query for "dark comedy" may not effectively distinguish between comedy books and comedy films if the genre vocabulary overlaps heavily. Third, the framework is non-personalized: it does not model individual user preferences, returning the same recommendations for all users submitting the same query. Fourth, the rating scale heterogeneity across domains (0–10 for books, 0.5–5.0 for movies, 2.3–5.0 for mobiles) means that raw hybrid scores are not directly comparable across domains, which is partially mitigated by the differential weighting but not fully resolved.

6.3 Comparison with Alternative Approaches

Compared to the deep learning-based cross-domain systems (e.g., BERT-based semantic search [13], neural collaborative filtering [12]), the proposed architecture trades representational richness for computational efficiency and interpretability. TF-IDF retrieval on sparse matrices is orders of magnitude faster than dense neural retrieval, making it suitable for real-time applications without GPU infrastructure. As Compared to pure collaborative filtering approaches, the content-based TF-IDF component handles the cold-start scenarios naturally, as new items with textual metadata can be immediately indexed and retrieved without requiring any rating history.

7. Conclusion

The unified cross-domain recommendation architecture, which operates over three domains, books, movies, and mobile devices, using TF-IDF embeddings, cosine similarity retrieval, and domain-adaptive hybrid fusion scoring. The architecture was first evaluated on three large-scale real-world datasets, containing over 84,803 items and across over 1.25 million user reviews. The key contributions in the group are the following: 1. A joint 5000-dimensional TF-IDF embedding space constructed from 33,114 items that enables the cross-domain semantic matching. 2. Domain-adaptive hybrid fusion scoring with the parameters alpha equal to 0.6, beta equal to 0.4, and gamma equal to 0.2, which fine-tunes against the actual dataset statistics. 3. A keyword-based domain inference model, which supports both targeted and exploratory retrieval. 4. A comprehensive empirical analysis that demonstrates the effectiveness of knowledge transfer. Experimental results confirm that the hybrid scoring mechanism effectively balances semantic relevance with quality stability. The hybrid retrieval achieves a top book score of 2.240 and a top movie score of 2.041. The Samsung Galaxy QD currently rules mobile retrieval with a top score of 0.558 at a cosine similarity of 0.650. Cross-dimension actually occurs by junior vocabulary bridging between the books and movies in the joint TF-IDF effective space.

Future work will explore that replaces TF-IDF with transformer-based embeddings (Sentence-BERT [14]) for richer semantic representations, that incorporates user interaction histories for personalized cross-domain recommendations systems, that developes learned weight optimization for fusion parameters, and expanding to additional domains such as music, fashion, and food domains. Formal evaluation using Precision@K, NDCG, and MAP metrics on held-out test sets will provide quantitative performance benchmarks.

Reference

1. Ziegler C-N, McNee SM, Konstan JA, Lausen G. Improving recommendation lists through topic diversification [Homepage on the Internet]. In: Proceedings of the 14th international conference on World Wide Web - WWW ’05. Chiba, Japan: ACM Press, 2005 [cited 2026 June 14]; p. 22.Available from: http://portal.acm.org/citation.cfm?doid=1060745.1060754
2. Harper FM, Konstan JA. The MovieLens Datasets: History and Context. ACM Trans Interact Intell Syst [homepage on the Internet] 2016 [cited 2026 June 14];5(4):1–19. Available from: https://dl.acm.org/doi/10.1145/2827872
3. Checking your browser - reCAPTCHA [Homepage on the Internet]. [cited 2026 June 14];Available from: https://www.kaggle.com/datasets/
4. Salton G, McGill MJ. Introduction to modern information retrieval. 3. pr. New York: McGraw-Hill Book Comp, 1987;
5. [Cited 2026 June 14]; Available from: https://www.semanticscholar.org/paper/Using-TF-IDF-to-Determine-Word-Relevance-in-Queries-Ramos/b3bf6373ff41a115197cb5b30e57830c16130c2c
6. Ricci F, Rokach L, Shapira B, editors. Recommender Systems Handbook [Homepage on the Internet]. Boston, MA: Springer US, 2015 [cited 2026 June 14]; Available from: https://link.springer.com/10.1007/978-1-4899-7637-6
7. Cantador I, Cremonesi P. Tutorial on cross-domain recommender systems [Homepage on the Internet]. In: Proceedings of the 8th ACM Conference on Recommender systems. Foster City, Silicon Valley California USA: ACM, 2014 [cited 2026 June 14]; p. 401–402.Available from: https://dl.acm.org/doi/10.1145/2645710.2645777
8. Pan SJ, Yang Q. A Survey on Transfer Learning. IEEE Trans Knowl Data Eng [homepage on the Internet] 2010 [cited 2026 June 14];22(10):1345–1359. Available from: http://ieeexplore.ieee.org/document/5288526/

9. Zhang Y, Ai Q, Chen X, Croft WB. Joint Representation Learning for Top-N Recommendation with Heterogeneous Information Sources [Homepage on the Internet]. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. Singapore Singapore: ACM, 2017 [cited 2026 June 14]; p. 1449–1458.Available from: https://dl.acm.org/doi/10.1145/3132847.3132892
10. Pazzani MJ, Billsus D. Content-Based Recommendation Systems [Homepage on the Internet]. In: Brusilovsky P, Kobsa A, Nejdl W, editors. The Adaptive Web. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007 [cited 2026 June 14]; p. 325–341.Available from: http://link.springer.com/10.1007/978-3-540-72079-9_10
11. Koren Y, Bell R, Volinsky C. Matrix Factorization Techniques for Recommender Systems. Computer [homepage on the Internet] 2009 [cited 2026 June 14];42(8):30–37. Available from: http://ieeexplore.ieee.org/document/5197422/
12. He X, Liao L, Zhang H, Nie L, Hu X, Chua T-S. Neural Collaborative Filtering [Homepage on the Internet]. In: Proceedings of the 26th International Conference on World Wide Web. Perth Australia: International World Wide Web Conferences Steering Committee, 2017 [cited 2026 June 14]; p. 173–182.Available from: https://dl.acm.org/doi/10.1145/3038912.3052569
13. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [Homepage on the Internet]. 2018 [cited 2026 June 14];Available from: https://arxiv.org/abs/1810.04805
14. Reimers N, Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks [Homepage on the Internet]. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics, 2019 [cited 2026 June 14]; p. 3980–3990.Available from: https://www.aclweb.org/anthology/D19-1410
15. Lops P, De Gemmis M, Semeraro G. Content-based Recommender Systems: State of the Art and Trends [Homepage on the Internet]. In: Ricci F, Rokach L, Shapira B, Kantor PB, editors. Recommender Systems Handbook. Boston, MA: Springer US, 2011 [cited 2026 June 14]; p. 73–105.Available from: https://link.springer.com/10.1007/978-0-387-85820-3_3
16. Burke R. Hybrid Recommender Systems: Survey and Experiments. User Model User-Adap Inter [homepage on the Internet] 2002 [cited 2026 June 14];12(4):331–370. Available from: https://link.springer.com/10.1023/A:1021240730564
17. Garreta R, Moncecchi G, editors. Learning scikit-learn: machine learning in Python: experience the benefits of machine learning techniques by applying them to real-world problems using Python and the open source scikit-learn library. Birmingham, UK: Packt Publishing Ltd, 2013;
18. Manning CD, Raghavan P, Schütze H. Introduction to information retrieval. Cambridge: Cambridge university press, 2008;