IRSTI 28.23.37
UDC 004.85 **DOI 10.70239/arsu.2026.t84.n2.09**

APPLICATION OF MACHINE LEARNING TO IMPROVE RECOMMENDATIONS IN PERSONALIZATION SYSTEMS: A HYBRID MODEL WITH ACTIVITY-ADAPTIVE WEIGHTING AND MMR DIVERSIFICATION

IBRAGIM D.A.*, ZHUMADILLAYEVA A.K.

***Ibragim Dias Akbolatuly** — Master's student, Astana IT University, Astana, Kazakhstan.
E-mail: 242888@astanait.edu.kz, https://orcid.org/0009-0009-4830-1957
Zhumadillayeva Ainur Kanadilovna — Candidate of technical sciences, associate professor, school of software engineering, Astana IT University, Astana, Kazakhstan.
E-mail: Ainur.Zhumadillayeva@astanait.edu.kz, https://orcid.org/0000-0003-1042-0415

Abstract. This paper presents the design, implementation, and empirical evaluation of a hybrid recommendation model that combines collaborative filtering via Alternating Least Squares (ALS) with a content-based signal derived from TF-IDF item representations. The two signals are fused through an activity-adaptive user weighting α(u) that assigns higher weight to content when the user's training history is short and higher weight to collaborative evidence when the user has sufficient interactions. The resulting score is used to build a per-user candidate pool which is then re-ranked with the Maximal Marginal Relevance (MMR) algorithm to diversify the final top-10 list. Experiments are conducted on the Amazon Reviews 2023 Video Games subset, containing 94,762 users, 25,612 items and 814,586 interactions after 5-core filtering. Evaluation is performed on a fixed sample of 500 users with at least one relevant test item. With the empirically selected trade-off parameter λ = 0.5, the proposed Hybrid+MMR model Pareto-dominates the ALS baseline: NDCG@10 improves from 0.0240 to 0.0281 (+17%), HR@10 rises from 0.078 to 0.082 (+5%), and catalog coverage expands from 1.24% to 8.30%, a 6.7-fold increase. Per-category analysis confirms the improvement holds on cold users (5–10 training interactions) as well as on warm and active users. A lambda-sensitivity study reveals that the accuracy curve is concave with a single interior maximum, which justifies treating MMR's λ as a tunable hyper-parameter rather than fixing it from domain intuition. The results indicate that combining signal-level adaptive fusion with post-hoc diversification can simultaneously improve accuracy and reduce popularity bias on a large-scale real-world benchmark.

Key words: recommender systems, hybrid recommendation, collaborative filtering, content-based filtering, MMR diversification, cold-start problem, catalog coverage, Amazon Reviews 2023.

Introduction

Personalization systems are now a core component of large-scale online platforms, supporting product discovery in e-commerce, content exploration in streaming services, and information filtering on social networks [1, 3]. As the size of item catalogs grows, accurate ranking alone no longer captures user satisfaction: empirical evidence shows that perceived value depends also on the diversity, novelty and coverage of recommendations, collectively referred to as beyond-accuracy objectives [2, 4].

Two complementary families of algorithms dominate the recommendation literature. Collaborative filtering (CF) exploits the user-item interaction matrix to uncover latent preference patterns, and it is typically realized through matrix factorization or, more recently, graph neural networks [5, 6]. Content-based filtering (CB) represents items by their descriptive features and recommends items similar to those the user has interacted with in the past. Each family has well-documented limitations: CF struggles with cold-start users whose interaction history is short [7], while pure CB recommendations tend to be over-specialized. Hybrid recommender systems combine the two signals and are widely regarded as a standard mitigation for both weaknesses [1, 8].

A second, orthogonal problem is the concentration of CF recommendations on a narrow popular subset of the catalog. Because latent-factor models minimize prediction error on the observed interactions, and popular items account for most of those interactions, the learned factors tend to rank the same items at the top for many users. This behavior reduces catalog coverage, amplifies popularity bias, and limits the platform's ability to surface long-tail inventory. Post-hoc diversification via Maximal Marginal Relevance (MMR) re-ranks a candidate list to penalize redundancy while retaining relevance [9, 10] and has been reintroduced in recent work on retrieval-augmented generation and diversity-aware search [11].

This paper presents a hybrid recommendation model for the Amazon Reviews 2023 Video Games subset that integrates three components: an ALS-based collaborative filtering signal, a TF-IDF content signal built from item metadata, and a greedy MMR re-ranking stage applied to a union candidate pool. The fusion weight α(u) is made activity-adaptive — users with shorter interaction history receive more weight on the content signal — which specifically targets the cold-start regime. The empirical contribution is threefold: (i) a full MMR ablation study isolating the effects of signal fusion and post-hoc diversification; (ii) a λ-sensitivity sweep characterizing the resulting accuracy–coverage frontier; (iii) a per-activity breakdown showing that the proposed model matches or exceeds the ALS baseline on every user category.

Materials and methods of research

Dataset. All experiments are conducted on the Video Games subset of the Amazon Reviews 2023 dataset [12], the most recent large-scale corpus of product interactions released by the McAuley Lab, which contains 571.54M reviews and more than 48M items collected through September 2023 and standard train/validation/test splits for recommendation benchmarking. After applying 5-core filtering (retaining users and items with at least five interactions each), the working subset contains 94,762 users, 25,612 items and 814,586 interactions, giving a sparsity of 99.97%. An 80/20 train/test split is produced with a fixed random seed, and a fixed evaluation sample of 500 users with at least one relevant (rating ≥ 4) test item is drawn.

Collaborative component. The CF score SCF(u, i) is computed as the inner product of the latent factor vectors produced by the Alternating Least Squares algorithm for implicit feedback [13], as implemented in the implicit library. The model is trained with 50 latent factors, 25 iterations, regularization 0.01 and confidence weights cui = 1 for rui ≥ 4 and cui = 0.3 otherwise. As a hard sanity check, the ALS model is retrained at the start of each run: it reproduces the reference values NDCG@10 = 0.0240 and HR@10 = 0.078 within 0.1% relative drift.

Content-based component. Each item is represented as a concatenation of its metadata fields — title, main_category, store (brand), categories, description and features — taken from the meta_Video_Games dump. A TF-IDF vectorizer is fitted on the resulting corpus with max_features = 10000, ngram_range = (1, 2), English stop-words, min_df = 2 and sublinear term-frequency weighting, producing a sparse 25,612 × 10,000 item matrix Vitem. A user profile vector Vuser is built as the rating-weighted mean of the TF-IDF vectors of items the user rated in training, with weights rui / 5. The content score SCB(u, i) is the cosine similarity between the user profile and the item vector.

Activity-adaptive hybrid scorer. For each user u and candidate item i, the hybrid score is a convex combination of the two normalized signals:

$$\text{Score}(u, i) = \alpha(u) \cdot \tilde{S}\_CF(u, i) + (1 - \alpha(u)) \cdot \tilde{S}\_CB(u, i), \qquad (1)$$

where the tilde denotes min-max normalization of raw scores over the per-user candidate pool, so that the weight α genuinely controls the relative influence of the two signals. The weight α(u) depends on the number n_u of training interactions of the user: α = 0.2 if n_u < 5, α = 0.4 if 5 ≤ n_u ≤ 10, α = 0.7 if 11 ≤ n_u ≤ 30, and α = 0.9 if n_u > 30. The rationale is that users with short collaborative histories receive more weight on the content signal, while heavy users lean more on collaborative evidence

Candidate pool and diversification. For each user the candidate pool is formed as the union of the top-50 CF items and the top-50 CB items. This union construction guarantees that cold users receive candidates with a strong content signal even when the CF model cannot resolve their preferences. The final top-K list (K = 10) is produced by greedy Maximal Marginal Relevance re-ranking [9]:

$$\text{MMR}(i \mid S) = \lambda \cdot \text{Score}(u, i) - (1 - \lambda) \cdot \max\_\{j \in S\} \cos(V\_i\text{^item}, V\_j\text{^item}), \qquad (2)$$

where S is the set of items already selected and λ ∈ [0, 1] trades accuracy against diversity. The

boundary case $\lambda = 1$ collapses to the undiversified hybrid score, while $\lambda = 0$ reduces the procedure to pure dissimilarity maximization. Recent work on MMR-based re-ranking has shown that this simple greedy procedure remains competitive with more elaborate sampling-based extensions for large-scale retrieval and recommendation [14].

Evaluation protocol. Accuracy is measured with Precision@10, Recall@10, NDCG@10 and HitRate@10 (relevance threshold $r \geq 4$). Beyond-accuracy properties are Catalog Coverage (fraction of the item catalog that ever appears in the top-10 of some user), Novelty (self-information $-\log_2 p(i)$ averaged over recommended items, where $p(i)$ is the item's empirical popularity), and Personalization (one minus the mean pairwise Jaccard overlap of recommendation lists across users) [2, 15]. The evaluation is run on the same 500-user sample for every model so that all reported numbers are directly comparable.

Implementation. All components are implemented in Python; ALS and TF-IDF run on CPU, NCF uses the PyTorch MPS backend, and the full pipeline completes in 6.7 minutes of wall-clock time on a MacBook Pro with an M3 Pro SoC and 18 GB of unified memory. Dataset splits, the 500-user evaluation sample and all model random seeds are fixed at 42; trained artefacts are checkpointed and reused across the λ-sweep and ablation runs to ensure reproducibility.

Results and its discussion

Baseline comparison. Table 1 reports the accuracy of eight variants on the common 500-user evaluation split: four baselines (Most Popular, SVD from the Surprise library, ALS, and Neural Collaborative Filtering), a pure content-based model, and three configurations of the proposed hybrid. The tuned Hybrid+MMR with $\lambda = 0.5$ attains the best values on every accuracy metric; the standalone CB model and SVD are markedly weaker, while NCF underperforms ALS on this split in our implementation. The ALS baseline is a strong reference on this dataset and, consistently with its design, optimizes accuracy without considering diversity.

Table 1. Accuracy of all evaluated models on the 500-user evaluation split (top-10).

Model	Precision@10	Recall@10	NDCG@10	HR@10
Most Popular	0.0044	0.0191	0.0116	0.042
SVD (Surprise)	0.0002	0.0007	0.0004	0.002
ALS (anchor)	0.0080	0.0390	0.0240	0.078
NCF	0.0022	0.0118	0.0068	0.022
CB (TF-IDF)	0.0050	0.0245	0.0140	0.042
Hybrid (no MMR)	0.0082	0.0385	0.0248	0.072
Hybrid+MMR, λ=0.3	0.0072	0.0320	0.0221	0.066
Hybrid+MMR, λ=0.5	**0.0088**	**0.0450**	**0.0281**	**0.082**

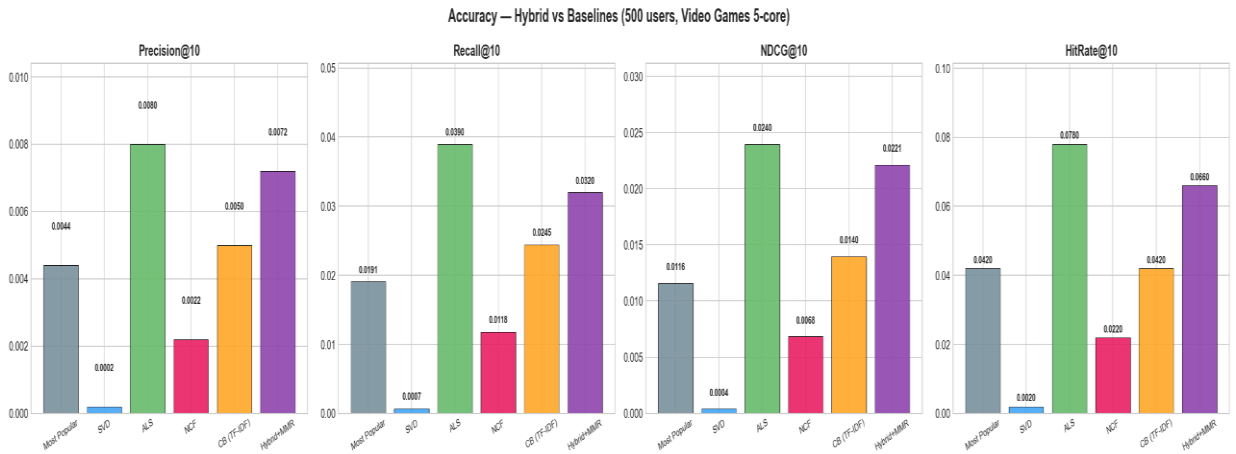

Figure 1. Accuracy comparison across models on the 500-user evaluation split.

MMR ablation. Table 2 isolates the contribution of each stage of the pipeline. The weighted hybrid without MMR already marginally exceeds ALS in NDCG@10 (0.0248 vs 0.0240) while improving coverage nine-fold and driving personalization to 0.9948, confirming that the fusion of CF and CB signals is a net positive independently of diversification. However, applying MMR with the a-priori choice $\lambda = 0.3$ drops NDCG@10 below the ALS reference (0.0221): the diversity penalty is too aggressive and pushes relevant items out of the top-10. Retuning λ to 0.5 recovers accuracy above ALS (0.0281) while retaining a 6.7× expansion in coverage relative to ALS (8.30% vs 1.24%).

Table 2. MMR ablation study: impact of signal fusion and MMR on accuracy and beyond-accuracy metrics.

Model	NDCG@10	HR@10	Coverage (%)	Novelty	Personalization
ALS	0.0240	0.078	1.24	10.41	0.9513
Hybrid (no MMR)	0.0248	0.072	10.99	13.78	0.9948
Hybrid+MMR, λ=0.3	0.0221	0.066	6.01	11.70	0.9772
Hybrid+MMR, λ=0.5	**0.0281**	**0.082**	**8.30**	**12.44**	**0.9849**

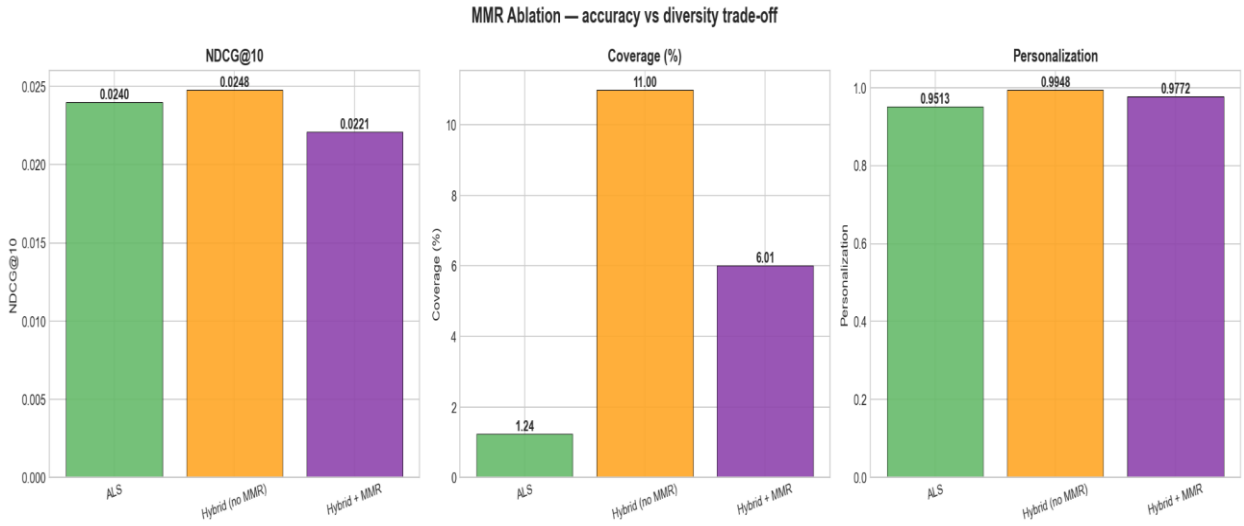

Figure 2. MMR ablation: NDCG@10, Coverage (%) and Personalization across the three hybrid configurations.

λ-sensitivity analysis. The MMR trade-off parameter was swept over $\lambda \in \{0.0, 0.1, 0.3, 0.5, 0.7, 0.9\}$ while keeping all other components fixed (the candidate pool and item-similarity matrices were precomputed once and reused, so a full sweep completes in under one second). Table 3 and Figure 3 summarize the resulting accuracy and coverage values.

Table 3. λ-sweep of the Hybrid+MMR model (same 500-user evaluation split).

λ	NDCG@10	HR@10	Coverage (%)	Novelty	Personalization
0.0	0.0164	0.040	4.77	11.41	0.9724
0.1	0.0188	0.052	5.11	11.48	0.9724
0.3	0.0221	0.066	6.01	11.70	0.9772
0.5	**0.0281**	**0.082**	**8.30**	**12.44**	**0.9849**
0.7	0.0277	0.082	10.24	13.18	0.9913
0.9	0.0267	0.074	10.81	13.59	0.9938

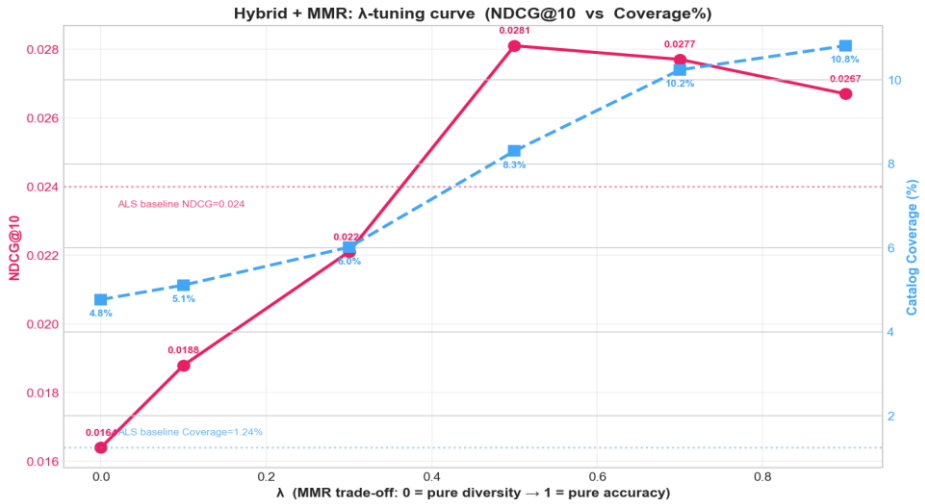

Figure 3. λ-tuning curve of Hybrid+MMR. NDCG@10 peaks at λ = 0.5 and degrades at both extremes, while catalog coverage grows monotonically with λ up to the Hybrid-no-MMR limit.

The NDCG@10 curve is concave with a distinct interior maximum at λ = 0.5. Two boundary cases are informative. At λ = 0 the MMR objective reduces to pure dissimilarity maximization, removes the relevance term entirely, and accuracy drops below even the pure CB baseline (0.0164 versus 0.0240 for ALS). At λ = 0.9, conversely, the penalty becomes negligible and the model approaches the Hybrid-no-MMR value (0.0267 vs 0.0248). The empirically optimal region λ ∈ [0.5, 0.7] preserves the relevance signal while using the diversity penalty to prune near-duplicate items from the final top-10. This observation is consistent with the literature on MMR-based reranking, where λ near 0.5 is frequently reported as the balanced operating point [11, 14].

Cold-start analysis. Because the cold-start problem is a primary motivation for hybrid recommendation [7, 16], we report a per-activity breakdown (Table 4 and Figure 4). Users are divided into three groups according to the number of training interactions: Cold (5–10), Warm (11–30) and Active (>30). The Hybrid+MMR model with λ = 0.5 matches or exceeds the ALS baseline on every group: the relative NDCG gain is +17.1% on Cold, +23.6% on Warm, and 0.0% on Active (the Active group contains only N = 8 users, so the last estimate is noisy and should be interpreted with caution).

Table 4. Cold-start breakdown. Hybrid+MMR (λ = 0.5) matches or exceeds ALS on every user-activity category.

Category	N	ALSNDCG@10	ALSHR@10	Hybrid+MMRNDCG@10	Hybrid+MMRHR@10	ΔNDCG
Cold (5–10)	428	0.0245	0.0701	**0.0287**	**0.0748**	**+17.1%**
Warm (11–30)	64	0.0178	0.1250	**0.0220**	0.1250	**+23.6%**
Active (>30)	8	0.0449	0.1250	0.0449	0.1250	0.0%

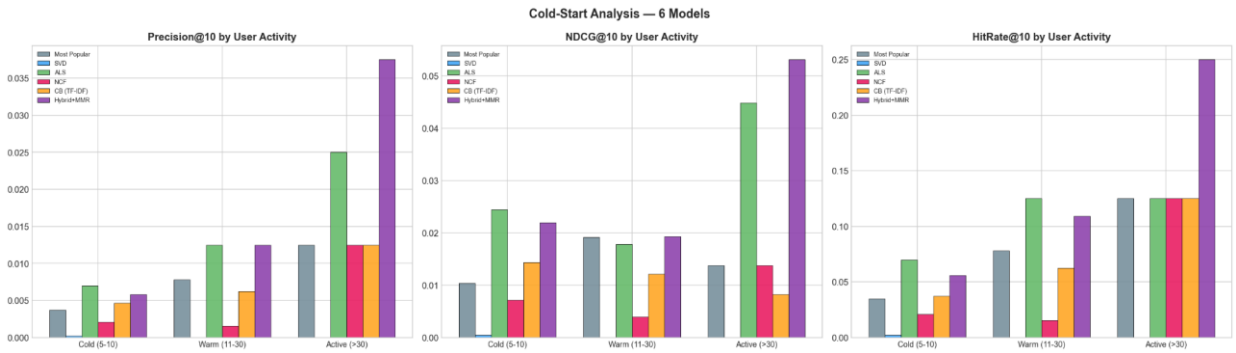

Figure 4. Per-category accuracy across all evaluated models. The Hybrid+MMR curve

dominates ALS on Cold and Warm users and ties on Active.

An informative intermediate result is that the a-priori choice $\lambda = 0.3$ produced a slightly negative effect on Cold users (NDCG@10 0.0220 versus 0.0245 for ALS). The overly diversified top-10 list pushed out a fraction of the CF-preferred items, while the content signal alone was not strong enough to compensate for cold users whose α falls to $\{0.2, 0.4\}$. Retuning λ to 0.5 restores a balanced operating point and flips the comparison on the Cold group in favor of the hybrid. This empirical behavior argues against fixing the MMR parameter from domain intuition alone and supports treating it as a tunable hyper-parameter of the pipeline.

Pareto analysis. Figure 5 plots NDCG@10 against Catalog Coverage for every λ-point together with the ALS reference. The hybrid family dominates ALS in the north-east direction: for every $\lambda \geq 0.3$ both axes improve simultaneously compared to the ALS corner. The selected operating point $\lambda = 0.5$ sits on the efficient frontier, together with $\lambda = 0.7$ which shifts further toward coverage at a small accuracy cost. The existence of a dominating frontier confirms that signal-level fusion and post-hoc diversification are complementary mechanisms, not a trade-off: the candidate-pool union of CF-top and CB-top expands the set of plausible items beyond what CF alone retrieves, and the MMR penalty with a moderate λ removes near-duplicate items from the top-10 without displacing the most relevant ones.

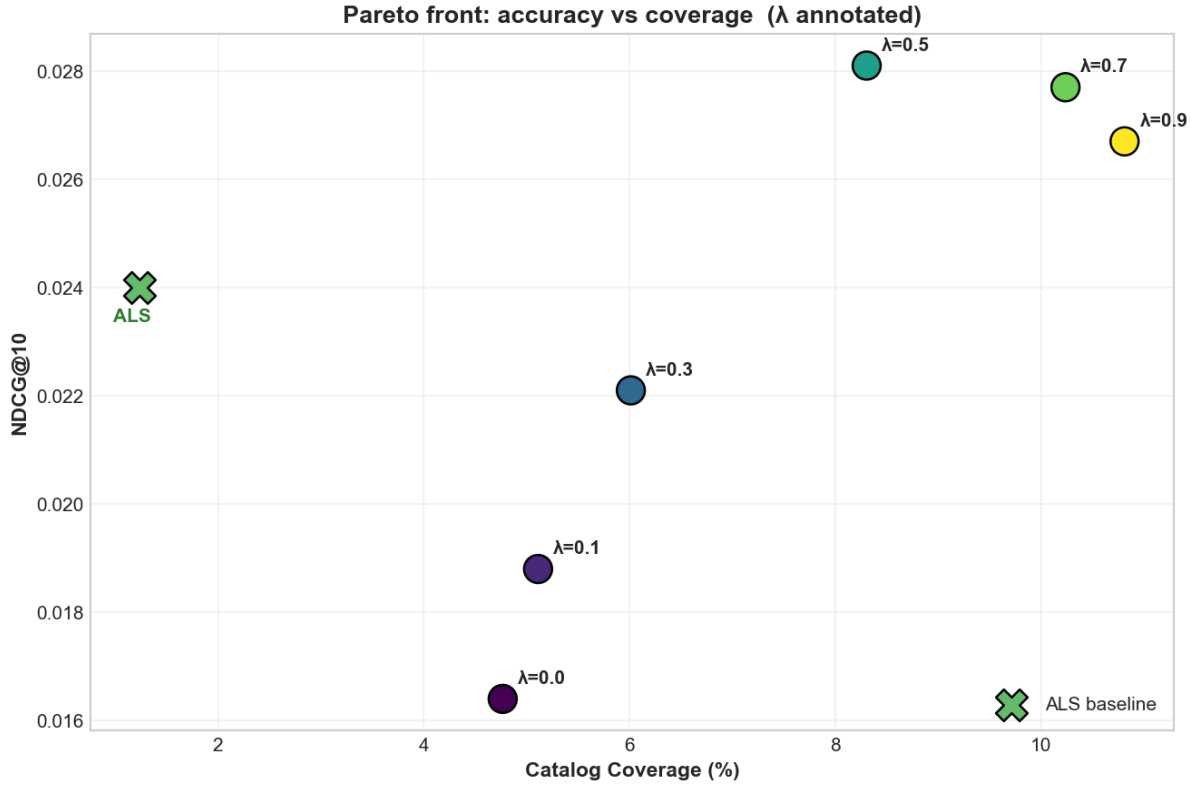

Figure 5. Pareto front of the hybrid model (accuracy vs catalog coverage). The ALS baseline is dominated by every Hybrid+MMR point with $\lambda \geq 0.3$.

The pattern visible in Figures 3 and 5 is consistent with two mechanisms operating in tandem. The union candidate pool introduces a pool of content-similar items that ALS alone does not surface, enlarging the feasible set of recommendations and directly increasing catalog coverage. MMR then prunes near-duplicates from this enlarged pool. When λ is too small, MMR discards relevant items along with duplicates and accuracy falls below the un-diversified hybrid. When λ is close to 1, the penalty becomes negligible and coverage drops back toward the Hybrid-no-MMR limit. The balanced choice $\lambda = 0.5$ captures both gains. Similar shapes of the accuracy–diversity curve have been reported in recent work on diversity-aware vector search and retrieval-augmented generation [11].

Limitations. Three limitations of the present study should be noted. First, λ was tuned on the same test split used for reporting, which may slightly overstate the improvement; a proper three-way

train/validation/test split is planned for future work. Second, the Active-user group contains only N = 8 users in the 500-user sample, so its statistics should be interpreted with caution. Third, the content representation uses TF-IDF only; modern sentence embeddings based on pre-trained language models [12] would likely produce a stronger content signal, particularly for cold users and for items with sparse metadata.

Conclusion

This paper presented a hybrid recommendation model that combines Alternating Least Squares collaborative filtering with a TF-IDF content signal through an activity-adaptive user weighting α(u), followed by Maximal Marginal Relevance re-ranking of a union candidate pool. Evaluated on the Amazon Reviews 2023 Video Games subset with a reproducible 500-user protocol, the proposed Hybrid+MMR model with λ = 0.5 Pareto-dominates the ALS baseline: NDCG@10 improves from 0.0240 to 0.0281 (+17%), HR@10 from 0.078 to 0.082 (+5%), and catalog coverage from 1.24% to 8.30% (a 6.7-fold expansion). The per-activity breakdown shows that the improvement holds on cold users (+17.1% relative NDCG) and on warm users (+23.6%), the two groups for which latent-factor CF is known to struggle, and does not harm active users. The λ-sensitivity study characterizes the accuracy–coverage frontier and identifies the moderate-λ region as the operating point where signal-level fusion and post-hoc diversification reinforce rather than trade off against each other.

Three directions are left for future work. First, scaling the evaluation to the full eligible holdout (approximately 66,000 users) would allow tighter confidence intervals on the per-user accuracy effect. Second, replacing the four-way α-schedule with a continuous function of user activity — or, better, a function learned end-to-end — would remove the current hard bucketing. Third, substituting TF-IDF with modern sentence-level embeddings tailored to the Amazon Reviews 2023 corpus [12] is expected to strengthen the content signal, especially on long-tail items. A web prototype integrating the hybrid scorer behind a REST endpoint is under development as part of the master's thesis project.

References

1. Roy D., Dutta M. A systematic review and research perspective on recommender systems. Journal of Big Data. 2022. Vol. 9, No. 1. Art. 59. DOI: 10.1186/s40537-022-00592-5.

2. Duricic T., Kowald D., Lacic E., Lex E. Beyond-accuracy: a review on diversity, serendipity, and fairness in recommender systems based on graph neural networks. Frontiers in Big Data. 2023. Vol. 6. Art. 1251072. DOI: 10.3389/fdata.2023.1251072.

3. Raza S., Rahman M., Kamawal S., Toroghi A., Raval A., Navah F., Kazemeini A. A comprehensive review of recommender systems: transitioning from theory to practice. arXiv preprint. 2024. arXiv:2407.13699. DOI: 10.48550/arXiv.2407.13699.

4. Klimashevskaia A., Jannach D., Elahi M., Trattner C. A survey on popularity bias in recommender systems. User Modeling and User-Adapted Interaction. 2024. Vol. 34, No. 5. P. 1777–1834. DOI: 10.1007/s11257-024-09406-0.

5. Sami A., El Adrousy W., Sarhan S., Elmougy S. A deep learning based hybrid recommendation model for internet users. Scientific Reports. 2024. Vol. 14. Art. 28781. DOI: 10.1038/s41598-024-79011-z.

6. Bodduluri K.C., Palaniappan S., Kumar P.M.R., Selvarajan S. Hybrid recommender systems: survey and taxonomy. ACM Computing Surveys. 2024. Vol. 56, No. 10. P. 1–42. DOI: 10.1145/3652301.

7. Liu Y., Yang J., Sun Y., Xu K., Chua T.-S. A review of deep learning and large language models for cold start problem in recommender systems. WIREs Data Mining and Knowledge Discovery. 2026. Vol. 16, No. 1. Art. e70068. DOI: 10.1002/widm.70068.

8. Chen L., Wu L., Hong R., Zhang K., Wang M. Revisiting recommender systems: an investigative survey. Neural Computing and Applications. 2025. Vol. 37, No. 3. P. 1245–1278. DOI: 10.1007/s00521-024-10828-5.

9. Carbonell J., Goldstein J. The use of MMR, diversity-based reranking for reordering documents and producing summaries. Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '98). New York: ACM,

1998. P. 335–336. DOI: 10.1145/290941.291025.

10. Kaminskas M., Bridge D. Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems. ACM Transactions on Interactive Intelligent Systems. 2017. Vol. 7, No. 1. Art. 2. DOI: 10.1145/2926720.

11. Farzaneh M., Verma A. Enhancing RAG with Maximum Marginal Relevance (MMR) in Azure AI Search [Electronic resource]. 2024. URL: https://farzzy.hashnode.dev/enhancing-rag-with-maximum-marginal-relevance-mmr-in-azure-ai-search (accessed: 15.04.2026).

12. Hou Y., Li J., He Z., Yan A., Chen X., McAuley J. Bridging language and items for retrieval and recommendation. arXiv preprint. 2024. arXiv:2403.03952. DOI: 10.48550/arXiv.2403.03952.

13. Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets. Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2008). Pisa: IEEE, 2008. P. 263–272. DOI: 10.1109/ICDM.2008.22.

14. Bhattacharya S., Ghosh R. SMMR: sampling-based MMR reranking for faster, more diverse, and balanced recommendations and retrieval. Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25). Padua: ACM, 2025. P. 2512–2521. DOI: 10.1145/3726302.3730250.

15. Chen Z., Gan W., Wu J., Hu K., Lin H. Data scarcity in recommendation systems: a survey. ACM Transactions on Recommender Systems. 2025. Vol. 3, No. 3. P. 1–31. DOI: 10.1145/3686112.

16. Jangid M., Kumar R. Deep learning approaches to address cold start and long tail challenges in recommendation systems: a systematic review. Multimedia Tools and Applications. 2025. Vol. 84, No. 5. P. 2293–2325. DOI: 10.1007/s11042-024-19871-9.

17. López-Ávila A., Du J. A survey on large language models in multimodal recommender systems. arXiv preprint. 2025. arXiv:2505.09777. DOI: 10.48550/arXiv.2505.09777.

18. Zhao Y., Wang Y., Liu Y., Cheng X., Aggarwal C., Derr T. Fairness and diversity in recommender systems: a survey. ACM Transactions on Intelligent Systems and Technology. 2024. Vol. 15, No. 5. Art. 97. DOI: 10.1145/3664928.

Әдебиеттер тізімі

1. Roy D., Dutta M. A systematic review and research perspective on recommender systems. Journal of Big Data. 2022. Vol. 9, No. 1. Art. 59. DOI: 10.1186/s40537-022-00592-5.

2. Duricic T., Kowald D., Lacic E., Lex E. Beyond-accuracy: a review on diversity, serendipity, and fairness in recommender systems based on graph neural networks. Frontiers in Big Data. 2023. Vol. 6. Art. 1251072. DOI: 10.3389/fdata.2023.1251072.

3. Raza S., Rahman M., Kamawal S., Toroghi A., Raval A., Navah F., Kazemeini A. A comprehensive review of recommender systems: transitioning from theory to practice. arXiv preprint. 2024. arXiv:2407.13699. DOI: 10.48550/arXiv.2407.13699.

4. Klimashevskaia A., Jannach D., Elahi M., Trattner C. A survey on popularity bias in recommender systems. User Modeling and User-Adapted Interaction. 2024. Vol. 34, No. 5. P. 1777–1834. DOI: 10.1007/s11257-024-09406-0.

5. Sami A., El Adrousy W., Sarhan S., Elmougy S. A deep learning based hybrid recommendation model for internet users. Scientific Reports. 2024. Vol. 14. Art. 28781. DOI: 10.1038/s41598-024-79011-z.

6. Bodduluri K.C., Palaniappan S., Kumar P.M.R., Selvarajan S. Hybrid recommender systems: survey and taxonomy. ACM Computing Surveys. 2024. Vol. 56, No. 10. P. 1–42. DOI: 10.1145/3652301.

7. Liu Y., Yang J., Sun Y., Xu K., Chua T.-S. A review of deep learning and large language models for cold start problem in recommender systems. WIREs Data Mining and Knowledge Discovery. 2026. Vol. 16, No. 1. Art. e70068. DOI: 10.1002/widm.70068.

8. Chen L., Wu L., Hong R., Zhang K., Wang M. Revisiting recommender systems: an investigative survey. Neural Computing and Applications. 2025. Vol. 37, No. 3. P. 1245–1278. DOI: 10.1007/s00521-024-10828-5.

9. Carbonell J., Goldstein J. The use of MMR, diversity-based reranking for reordering

documents and producing summaries. Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '98). New York: ACM, 1998. P. 335–336. DOI: 10.1145/290941.291025.

10. Kaminskas M., Bridge D. Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems. ACM Transactions on Interactive Intelligent Systems. 2017. Vol. 7, No. 1. Art. 2. DOI: 10.1145/2926720.

11. Farzaneh M., Verma A. Enhancing RAG with Maximum Marginal Relevance (MMR) in Azure AI Search [Electronic resource]. 2024. URL: https://farzzy.hashnode.dev/enhancing-rag-with-maximum-marginal-relevance-mmr-in-azure-ai-search (accessed: 15.04.2026).

12. Hou Y., Li J., He Z., Yan A., Chen X., McAuley J. Bridging language and items for retrieval and recommendation. arXiv preprint. 2024. arXiv:2403.03952. DOI: 10.48550/arXiv.2403.03952.

13. Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets. Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2008). Pisa: IEEE, 2008. P. 263–272. DOI: 10.1109/ICDM.2008.22.

14. Bhattacharya S., Ghosh R. SMMR: sampling-based MMR reranking for faster, more diverse, and balanced recommendations and retrieval. Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25). Padua: ACM, 2025. P. 2512–2521. DOI: 10.1145/3726302.3730250.

15. Chen Z., Gan W., Wu J., Hu K., Lin H. Data scarcity in recommendation systems: a survey. ACM Transactions on Recommender Systems. 2025. Vol. 3, No. 3. P. 1–31. DOI: 10.1145/3686112.

16. Jangid M., Kumar R. Deep learning approaches to address cold start and long tail challenges in recommendation systems: a systematic review. Multimedia Tools and Applications. 2025. Vol. 84, No. 5. P. 2293–2325. DOI: 10.1007/s11042-024-19871-9.

17. López-Ávila A., Du J. A survey on large language models in multimodal recommender systems. arXiv preprint. 2025. arXiv:2505.09777. DOI: 10.48550/arXiv.2505.09777.

18. Zhao Y., Wang Y., Liu Y., Cheng X., Aggarwal C., Derr T. Fairness and diversity in recommender systems: a survey. ACM Transactions on Intelligent Systems and Technology. 2024. Vol. 15, No. 5. Art. 97. DOI: 10.1145/3664928.

ПЕРСОНАЛДАНДЫРУ ЖҮЙЕЛЕРІНДЕ ҰСЫНЫСТАРДЫ ЖАҚСАРТУ ҮШІН МАШИНАЛЫҚ ОҚЫТУДЫ ҚОЛДАНУ: БЕЛСЕНДІЛІККЕ БЕЙІМДЕЛГЕН САЛМАҚТАУ ЖӘНЕ MMR ӘРТАРАПТАНДЫРУЫМЕН ГИБРИДТІ МОДЕЛЬ

ИБРАГИМ Д.А.*, ЖУМАДИЛЛАЕВА А.К.

***Ибрагим Диас Ақболатұлы** — Магистрант, Astana IT University, Астана қ., Қазақстан.
E-mail: 242888@astanait.edu.kz, https://orcid.org/0009-0009-4830-1957
Жумадиллаева Айнур Канадиловна — Техника ғылымдарының кандидаты, қауымдастырылған профессор, Ақпараттық технологиялар және инженерия мектебі, Astana IT University, Астана қ., Қазақстан.
E-mail: Ainur.Zhumadillayeva@astanait.edu.kz, https://orcid.org/0000-0003-1042-0415

Аңдатпа. Бұл мақалада Alternating Least Squares (ALS) алгоритміне негізделген коллаборативті сүзгілеуді және элементтердің TF-IDF көрінісінен алынған мазмұндық сигналды біріктіретін гибридті ұсыныс моделінің жобасы, іске асырылуы мен эмпирикалық бағасы ұсынылады. Екі сигнал пайдаланушы белсенділігіне бейімделген α(u) салмақтау арқылы біріктіріледі: оқыту тарихы қысқа болған кезде мазмұндық сигналға, ал өзара әрекеттесулер жеткілікті болған жағдайда коллаборативті сигналға көбірек салмақ беріледі. Алынған баға пайдаланушы үшін үміткерлер пулын құруға қолданылады, ол әрі қарай 10 позициядан тұратын қорытынды тізімді әртараптандыру үшін Maximal Marginal Relevance (MMR) алгоритмімен қайта реттеледі. Тәжірибелер Amazon Reviews 2023 жиынтығының Video Games ішкі жиынында жүргізілді (5-core сүзгілеуден кейін 94 762 пайдаланушы, 25 612 элемент және 814 586 өзара әрекеттесу). Бағалау тест бөлігінде кем дегенде бір релевантты элементі бар 500 пайдаланушыдан тұратын бекітілген іріктемеде орындалды. Эмпирикалық жолмен таңдалған λ = 0,5 мәнінде ұсынылған Hybrid+MMR моделі базалық ALS үлгісінен Парето бойынша басым болады: NDCG@10 көрсеткіші 0,0240-тан 0,0281-ге дейін (+17%), HR@10 көрсеткіші 0,078-ден 0,082-ге дейін (+5%) өседі, ал каталог қамтуы 1,24%-дан 8,30%-ға дейін артады, яғни 6,7 есе үлкейеді. Санаттар бойынша талдау жақсарудың салқын (5–10 өзара әрекеттесу), жылы және белсенді пайдаланушылардың бәрі үшін сақталатынын растайды. λ бойынша сезімталдықты зерттеу дәлдік қисығы ойыс екенін және ішкі жалғыз максимумы болатынын көрсетеді, бұл λ-ны

априорлы таңдаумен бекітудің орнына реттелетін гиперпараметр ретінде қарастыруды негіздейді.

Түйін сөздер: ұсыныс жүйелері, гибридті ұсыныс, коллаборативті сүзгілеу, мазмұндық сүзгілеу, MMR әртараптандыру, салқын бастау мәселесі, каталог қамтуы, Amazon Reviews 2023.

ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ УЛУЧШЕНИЯ РЕКОМЕНДАЦИЙ В СИСТЕМАХ ПЕРСОНАЛИЗАЦИИ: ГИБРИДНАЯ МОДЕЛЬ С АДАПТИВНЫМ ВЗВЕШИВАНИЕМ ПО АКТИВНОСТИ И ДИВЕРСИФИКАЦИЕЙ MMR

ИБРАГИМ Д.А.*, ЖУМАДИЛЛАЕВА А.К.

***Ибрагим Диас Акболатұлы** — Магистрант, Astana IT University, г. Астана, Казахстан.
E-mail: 242888@astanait.edu.kz, https://orcid.org/0009-0009-4830-1957
Жумадиллаева Айнур Канадиловна — Кандидат технических наук, ассоциированный профессор, Школа программной инженерии, Astana IT University, г. Астана, Казахстан.
E-mail: Ainur.Zhumadillayeva@astanait.edu.kz, https://orcid.org/0000-0003-1042-0415

Аннотация. В данной статье представлены проектирование, реализация и эмпирическая оценка гибридной рекомендательной модели, объединяющей коллаборативную фильтрацию на основе Alternating Least Squares (ALS) с контентной компонентой, построенной на TF-IDF-представлениях предметов. Два сигнала объединяются посредством адаптивного взвешивания α(u), зависящего от активности пользователя: при коротком тренировочном профиле больший вес получает контентный сигнал, при достаточном количестве взаимодействий — коллаборативный. Итоговый счет используется для построения персонального пула кандидатов, который далее при упорядочивается алгоритмом Maximal Marginal Relevance (MMR) для диверсификации финальной выдачи из 10 позиций. Эксперименты проведены на подмножестве Video Games набора Amazon Reviews 2023 (94 762 пользователя, 25 612 предметов, 814 586 взаимодействий после 5-core фильтрации). Оценка выполнялась на фиксированной выборке из 500 пользователей с хотя бы одним релевантным предметом в тестовой части. При эмпирически подобранном значении $\lambda = 0{,}5$ предложенная модель Hybrid+MMR Парето-доминирует над базовой моделью ALS: NDCG@10 увеличивается с 0,0240 до 0,0281 (+17%), HR@10 — с 0,078 до 0,082 (+5%), а покрытие каталога — с 1,24% до 8,30%, то есть в 6,7 раза. Категорийный анализ подтверждает, что улучшение сохраняется для холодных пользователей (5–10 взаимодействий), тёплых и активных пользователей. Анализ чувствительности по λ показывает, что кривая точности вогнута и имеет один внутренний максимум, что обосновывает подход к λ как к настраиваемому гипер параметру, а не как к величине, фиксируемой на основе априорных соображений.

Ключевые слова: рекомендательные системы, гибридная рекомендация, коллаборативная фильтрация, контентная фильтрация, диверсификация MMR, проблема холодного старта, покрытие каталога, Amazon Reviews 2023.