ISNN 3030-3907

2026
MAY
ISSUE - 5
VOLUME - 6

DEVELOPMENT OF SCIENCE

scientific journal



Development of science

Ilmiy jurnal

MAY 2026
Issue - 5, VOLUME – 6

ISSN 3030 -3907

Ilmiy jurnal OAK rayosatining 2024 yil 27 sentabrdagi 361-son qaroriga asosan 02.00.00 kimyo fanlari hamda 2025 yil 12 fevral 367- son qaroriga asosan 05.00.00 texnika fanlari bo‘yicha falsafa doktori (PhD) va fan doktori (DSc) ilmiy darajasiga talabgorlarning dissertatsiya ishlari yuzasidan dissertatsiyalari asosiy ilmiy natijalarini chop etish tavsiya etilgan milliy ilmiy nashrlar ro‘yxatiga kiritilgan.

BOSH MUHARRIR:
Temirov Alisher Hoshim o‘g‘li – texnika fanlari falsafa doktori (PhD),Dotsent.
TAHRIRIYAT KENGASH RAISI:
Fozilov Sadriddin Fayzullayevich – texnika fanlari doktori, professor
TAHRIRIYAT KENGASH A’ZOLARI

TEXNIKA FANLARI/ TECHNICAL SCIENCES/ ТЕХНИЧЕСКИЕ НАУКИ

Bozorov G‘ayrat Rashidovich
Buxoro Davlat Texnika Universiteti, professor
Abdurahmonov Olim Rustamovich
Buxoro Davlat Texnika Universiteti, professor
Adizov Bobur Zamirovich
O‘zbekiston Respublikasi Fanlar akademiyasi Umumiy va noorganik kimyo instituti, professor
Hayitov Ruslan Rustamjonovich
Buxoro Davlat Texnika Universiteti, professor
Jumayev Qayum Karimovich
Buxoro Davlat Texnika Universiteti, professor.
Berdiyev Dorob Murodovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor
Maxmudov Rafiq Amonovich
Buxoro Davlat Texnika Universiteti, professor.
Tilloyev Lochin Ismatilloyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Raxmonov Qahramon Sanakulovich
Buxoro Davlat Texnika Universiteti, dotsent.
Barakayev Nusratullo Rajabovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor.
Jo‘rayev Xayrullo Fayziyevich
Buxoro Davlat Texnika Universiteti, professor.
Panoyev Erali Rajabboyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Ataullayev Sherzod Nabiullyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Malik Negmurodovich
Buxoro Davlat Texnika Universiteti, dotsent.
Axmedova Ozoda Bahronovna
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti.
Safarov Bahri Jumayevich
Buxoro Davlat Texnika Universiteti, dotsent.
Xaydarov Axtam Amonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Safarov Jasur Alijon o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.

KIMYO FANLARI/ CHEMICAL SCIENCES/ ХИМИЧЕСКИЕ НАУКИ

Axmedov Voxid Nizomovich
Buxoro Davlat Texnika Universiteti, professor
Umarov Baqo Bafoyevich
Buxoro Davlat Universiteti, professor
Kodirov Abduaxad Abduraximovich
Qarshi davlat Universiteti, professor
Qurbonov Mingniqul Jumagulovich
Qarshi davlat Universiteti, dotsent
Turemuratov Sharibay Naurizbaevich
O‘zbekiston Respublikasi Fanlar akademiyasi Qoraqalpog‘iston bo‘limi Bosh ilmiy kotibi, professor
Tursunov Kaxor Shonazarovich
Qarshi davlat Universiteti, professor
Do‘stov Hamro
Buxoro Davlat Texnika Universiteti, professor
Panoyev Nodir Shavkatovich
Buxoro Davlat Texnika Universiteti, dotsent.
Olimov Bobur Bahodir o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.
Kasimov Sherzod Abduazimovich
Termiz Davlat Universiteti, professor.
Aliqulov Rustam Valiyevich
Termiz Davlat Universiteti, professor
Tropov Xamza Tursunovich
Samarqand Davlat Universiteti, professor

IQTISODIYOT FANLARI/ ECONOMICS/ ЭКОНОМИЧЕСКИЕ НАУКИ

Nizamov Aslitdin Badreddinovich
Buxoro Davlat Texnika Universiteti, professor.
Boboyev Akmal Choriyevich
Buxoro Davlat Texnika Universiteti, professor.
Ergashxojayeva Shaxnoza Djasurovna
Toshkent davlat iqtisodiyot universiteti, professor
Saidaxmedova Nodira Ilxomovna
Toshkent davlat iqtisodiyot universiteti, professor
Muxsinov Bekzod Toxirovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shakarov Qulmat Ashirovich
Renessans ta'lim universiteti,dotsent.
Jumaeva Zulfiya Qayumovna
Buxoro Davlat Texnika Universiteti, dotsent.
Norova Salomat Yusupovna
Buxoro Davlat Texnika Universiteti, dotsent.
Avezova Shaxnoza Maxmudjonovna
Buxoro Davlat Texnika Universiteti, dotsent.
Jalilov Jamshid G‘anijonovich
Toshkent davlat iqtisodiyot universiteti, dotsent
Xasanova Gulruh Djumanazarovna
Buxoro Davlat Texnika Universiteti, dotsent.

PEDAGOGIKA FANLARI/ PEDAGOGICAL SCIENCES / ПЕДАГОГИЧЕСКИЕ НАУКИ

Siddiqova Sadoqat G‘afforovna
Buxoro Davlat Texnika Universiteti, dotsent.
Zoyirov Erkin Xalilovich
Buxoro Davlat Texnika Universiteti, DSc, professor
Gaffarov Feruz Hasanovich
Puchon Universiteti, *professor*
Ruziyev Davron Yuldashevich
Buxoro Davlat Universiteti, professor
Muradova Firuza Rashidovna
Buxoro Davlat Texnika Universiteti, professor.
Yunusova Gulandon Samiyevna
Buxoro Davlat Texnika Universiteti, professor.
To‘xtayeva Zebo Sharifovna
Buxoro Davlat Texnika Universiteti, professor.
Navro‘zova Gulchixra Nigmatovna
Buxoro Davlat Texnika Universiteti, dotsent.
Uzoqova Laylo Polvonovna
Buxoro muhandislik – texnologiya instituti, dotsent.
To‘raqulova Marjona Qiyom qizi
Buxoro Davlat Texnika Universiteti, dotsent.
Muradova Zarina Rashidovna
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Sanjar Aslonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shodiyev Jahongir Jo‘raqulovich
Buxoro Davlat Texnika Universiteti, dotsent
Gaffarov Laziz Hasanovich
Buxoro muhandislik – texnologiya instituti, dotsent.

Tahririyat a’zolarning to‘liq ro‘yxati jurnalning www.devos.uz saytida keltirilgan.

Tahririyat hay’ati a’zolarining tarkibini o‘zgartirish hamda qo‘shimcha kiritish yoki tahririyat a’zoligidan chiqarish bosh muharrir tomonidan amalga oshiriladi.

ISSN 3030-3907

Development of science ilmiy jurnali 2024 yil 15 aprel kuni №256445 sonli guvohnoma bilan Davlat ro‘yxatidan o‘tgan.

To‘plam elektron ko‘rinishda (pdf formatida) mualliflarga taqdim etiladi. To‘plamga kiritiladigan maqolalarning mazmuni, ulardagi statistik ma’lumotlar va me’yoriy hujjatlarning to‘g‘riligiga hamda har qanday fikr - mulohazalarga mualliflarning o‘zlari mas’ul hisoblanadilar. Talablarga javob bermaydigan maqolalar to‘plamga kiritilmaydi. Tashkiliy qo‘mita maqola matnini qisqartirish, qisman tuzatish kiritish, seksiyalarga joylashtirish huquqiga ega.

ISSN 3030-3907

A HYBRID RECOMMENDER SYSTEM FOR ONLINE PLATFORMS USING COLLABORATIVE AND CONTENT-BASED FILTERING WITH BAYESIAN HYPERPARAMETER OPTIMIZATION

Muxtarova A.I
Muhammad Al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti magistranti
E-mail: azizamukhtarova26@gmail.com
Artikiva M.A
Muhammad Al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti magistranti

Abstract. This paper presents a hybrid recommender system designed for online platforms, combining Singular Value Decomposition (SVD)-based collaborative filtering with TF-IDF-based content-based filtering. The system addresses the core limitations of single-paradigm approaches - data sparsity and the cold-start problem - through a weighted combination mechanism whose parameters are jointly optimised using Bayesian search via the Optuna library. Experiments conducted on the Amazon Reviews 2023 Video Games dataset (686,004 interactions, 89,615 users, 23,531 items, 99.5% sparsity) demonstrate that the content-based component achieves Precision@10 = 0.0014 and MRR = 0.0038, outperforming the popularity baseline across personalisation metrics. A three-tier fallback strategy handles cold-start scenarios, and the system is deployed as a REST API on Hugging Face Spaces, demonstrating real-time applicability.

Keywords: recommender systems, collaborative filtering, content-based filtering, hybrid model, Bayesian optimisation, Optuna, cold-start problem, matrix factorization

Annotatsiya. Ushbu maqolada onlayn platformalar uchun mo'ljallangan gibrid tavsiya tizimi taqdim etiladi. Tizim SVD asosidagi kollaborativ filtrlash va TF-IDF asosidagi kontent filtrlashni birlashtiradi. Yagona paradigmali yondashuvlarning asosiy cheklovlari — ma'lumotlar siyrakligi va sovuq ishga tushirish muammosi — Optuna kutubxonasi orqali Bayes qidiruvi yordamida parametrlar bir vaqtda optimallashtirilgan og'irlikli birlashtirish mexanizmi orqali hal etiladi. Amazon Reviews 2023 Video Games dataseti (686 004 ta o'zaro ta'sir, 89 615 foydalanuvchi, 23 531 mahsulot, 99,5% siyraklik) ustida o'tkazilgan tajribalar kontent asosidagi komponentning Precision@10 = 0,0014 va MRR = 0,0038 ko'rsatkichlariga erishganini hamda shaxsiylashtirish metrikalari bo'yicha mashhurlik asosidagi modeldan ustun kelganini ko'rsatdi. Uch bosqichli fallback strategiyasi sovuq ishga tushirish holatlarini hal qiladi. Tizim Hugging Face Spaces platformasida REST API sifatida joylashtirilgan bo'lib, real vaqt rejimida qo'llanish imkoniyatini namoyish etadi.

ISSN 3030-3907

Kalit so'zlar: tavsiya tizimlari, kollaborativ filtrlash, kontent asosidagi filtrlash, gibrid model, Bayes optimizatsiyasi, Optuna, sovuq ishga tushirish muammosi, matritsa faktorizatsiyasi

Аннотация. В данной статье представлена гибридная рекомендательная система для онлайн-платформ, сочетающая коллаборативную фильтрацию на основе сингулярного разложения (SVD) с контентной фильтрацией на основе TF-IDF. Система устраняет основные ограничения подходов с единственной парадигмой — разреженность данных и проблему холодного старта — с помощью механизма взвешенного объединения, параметры которого совместно оптимизируются байесовским поиском через библиотеку Optuna. Эксперименты, проведённые на наборе данных Amazon Reviews 2023 категории Video Games (686 004 взаимодействия, 89 615 пользователей, 23 531 товар, разреженность 99,5%), показали, что контентный компонент достигает Precision@10 = 0,0014 и MRR = 0,0038, превосходя базовую модель на основе популярности по метрикам персонализации. Трёхуровневая стратегия резервного переключения обрабатывает сценарии холодного старта. Система развёрнута в виде REST API на платформе Hugging Face Spaces, демонстрируя применимость в режиме реального времени.

Ключевые слова: рекомендательные системы, коллаборативная фильтрация, контентная фильтрация, гибридная модель, байесовская оптимизация, Optuna, проблема холодного старта, матричная факторизация

1. INTRODUCTION.

The proliferation of digital content has created an *information overload* problem in which users struggle to identify relevant items within catalogues of millions of products. Recommender systems address this challenge by personalizing the information landscape for each user. Amazon attributes approximately 35% of its revenue to recommendations [1], and Netflix reports that over 80% of streamed content is discovered via its recommendation engine [2]. Despite their commercial importance, recommender systems face persistent technical challenges that limit performance in real-world deployments.

Two dominant paradigms exist in the literature. Collaborative filtering (CF) exploits patterns in user-item interaction history to identify users with similar preferences [3]. Content-based filtering (CB) represents items by their intrinsic attributes and builds user profiles from the characteristics of previously consumed items [4]. Each paradigm exhibits complementary weaknesses: CF suffers from data sparsity and the cold-start problem, while CB is prone to over-specialisation and depends on metadata quality. Hybrid systems that combine both paradigms have consistently demonstrated superior performance [5].

ISSN 3030-3907

The present work makes the following contributions. First, a hybrid weighted combination of SVD-based CF and TF-IDF-based CB is implemented in Python using modular, reusable components. Second, all system hyperparameters including the combination coefficient α, SVD latent factor count, and TF-IDF vocabulary size - are jointly optimised using Bayesian search, avoiding the sub-optimal sequential tuning prevalent in prior work. Third, a three-tier fallback mechanism is introduced to handle cold-start users without degrading the experience for active users. Fourth, the complete system is deployed as a REST API and evaluated on the publicly available Amazon Reviews 2023 dataset.

2. RELATED WORK

Matrix factorisation methods were established as the state of the art for collaborative filtering by Koren et al. [3] in the context of the Netflix Prize. Their SVD model decomposes the user-item interaction matrix R into latent factor matrices P and Q such that $R \approx P \cdot Q^T$, learning personalised representations through stochastic gradient descent. Extensions including SVD++ [3] and Neural Collaborative Filtering [6] have further improved accuracy by incorporating implicit feedback and non-linear interaction functions respectively.

Content-based approaches represent items using TF-IDF weighted term vectors or dense embeddings from pre-trained language models such as Sentence-BERT [7]. User profiles are constructed by aggregating the vectors of positively rated items. Cosine similarity between the profile vector and candidate item vectors drives recommendation ranking.

Hybrid architectures are classified by Burke [5] into seven categories, of which weighted combination is the most widely studied. Most published implementations treat the combination coefficient α as a static scalar tuned independently of the component model parameters. This sequential approach cannot capture parameter interactions across components. The present work addresses this gap by optimising α jointly with all other hyperparameters.

Bayesian hyperparameter optimisation using Tree-structured Parzen Estimators (TPE) [8] has been shown to outperform grid and random search for machine learning models. Its application to hybrid recommender systems with joint parameter spaces has received limited attention in the literature, motivating the approach adopted here.

3. METHODS.

3.1 Dataset and Preprocessing.

Experiments were conducted on the Amazon Reviews 2023 dataset [9], Video Games category. The raw dataset contains 4.6 million interactions. After applying 5-core filtering - retaining only users and items with at least five interactions - the working dataset comprises 686,004 interactions across 89,615

ISSN 3030-3907

users and 23,531 items, yielding a sparsity of 99.5%. Each interaction record contains a user identifier, item identifier, rating (1–5), review text, and timestamp. Ratings were normalised to [0, 1] using the formula $r_n = (r - 1)/4$. A temporal split was applied: for each user, interactions were sorted chronologically, with the first 80% assigned to training, the next 10% to validation, and the final 10% to test. This temporal partitioning simulates the production setting in which models are trained on historical data and evaluated on future behaviour.

3.2 Collaborative Filtering Component.

The CF component implements matrix factorisation with user and item biases in PyTorch. Each user u and item i are represented by k-dimensional latent factor vectors p_u and q_i together with scalar bias terms b_u and b_i. The predicted rating is given by:

$$\hat{r}(u, i) = \mu + b_u + b_i + p_u \cdot q_i^T$$

where μ denotes the global mean rating. Parameters are optimised with the Adam optimiser using a cosine annealing learning rate schedule over 100 epochs. Initial embedding weights are drawn from N(0, 0.1) to ensure non-zero gradient signal from the first iteration. The final configuration uses k = 150 latent factors and a learning rate of 0.01.

3.3 Content-Based Filtering Component.

Each item is represented by a TF-IDF vector computed over the concatenation of its title, features, and description fields. The vocabulary is limited to the top max_features terms by document frequency. A user profile vector is constructed as the rating-weighted mean of the TF-IDF vectors of items rated in the training set. Recommendations are generated by ranking candidate items by cosine similarity to the user profile, excluding items already rated.

3.4 Hybrid Combination and Optimisation.

Both component scores are min-max normalised to [0, 1] before combination. The hybrid score is computed as:

$$\text{Score}(u, i) = \alpha \cdot \text{CF}_{\text{norm}}(u, i) + (1 - \alpha) \cdot \text{CB}_{\text{norm}}(u, i)$$

Hyperparameters are jointly optimised using Optuna [8] with the TPE sampler. The search space encompasses n_factors ∈ [20, 200], α ∈ [0, 1], and max_features ∈ [500, 5000]. The objective is NDCG@10 on the validation set, evaluated over 50 trials. Unpromising trials are pruned using the Hyperband pruner to reduce total computation time.

3.5 Cold-Start Strategy.

A three-tier fallback mechanism addresses cold-start scenarios. Users absent from the training set receive popularity-based recommendations ranked by interaction count and mean rating. Users with 1-4 training interactions receive CB-dominant recommendations, since their interaction history is insufficient for

ISSN 3030-3907

reliable CF latent factor estimation. Users with five or more training interactions receive full hybrid recommendations with the optimised α. This progressive personalisation approach ensures the system is immediately useful for new users while delivering fully personalised recommendations to active users.

4. RESULTS.

All models were evaluated on the held-out test set using four ranking metrics: Precision@10, Recall@10, NDCG@10, and Mean Reciprocal Rank (MRR). Evaluation was performed over 500 randomly sampled users who appear in both training and test sets and have at least one reachable relevant item - defined as an item rated ≥ 4 stars that also appears in the training catalogue.

Table 1.

Model comparison on the test set (@K = 10)

Model	Precision@10	Recall@10	NDCG@10	MRR
Popularity Baseline	0.0012	0.0076	0.0033	0.0034
CF (SVD)	0.0000	0.0000	0.0000	0.0000
CB (TF-IDF)	0.0014	0.0050	0.0032	0.0038
Hybrid (Optuna)	0.0014	0.0050	0.0029	0.0037

The optimised parameters identified by Optuna are: n_factors = 45, α = 0.29, max_features = 2,000. The CB component thus receives 71% of the total weight in the hybrid score.

Several observations are noteworthy. First, both CB and Hybrid outperform the popularity baseline on Precision@10 and MRR, confirming that content signals provide personalisation value beyond global popularity. Second, the CF component achieves zero performance across all metrics. Debug analysis reveals that the SVD model consistently recommends items from the early 1990s and 2000s that do not appear in the test set - a manifestation of popularity bias in sparse interaction matrices, where item bias terms dominate the prediction and drive recommendations toward globally frequent items regardless of user-specific factors. Third, the Hybrid model achieves identical Precision@10 and Recall@10 to CB alone, which is expected given α = 0.29 and the non-contributing CF component: the hybrid effectively reduces to CB at this operating point. The marginal difference in NDCG@10 (0.0032 vs 0.0029) and MRR (0.0038 vs 0.0037) reflects the small weight assigned to the CF score distorting the ranking of a small number of items.

5. DISCUSSION.

ISSN 3030-3907

The principal finding of this study is that content-based signals derived from item metadata provide reliable personalisation even under extreme interaction sparsity (99.5%), whereas matrix factorisation collapses to a popularity-based ranking in this regime. This result aligns with the theoretical analysis of Linden et al. [1] and the empirical observations of Schein et al. [10], who note that latent factor models require sufficient interaction density to learn user-specific representations.

The failure of the CF component has a direct implication for the hybrid combination: Bayesian optimisation correctly identified $\alpha = 0.29$, assigning minimal weight to a non-contributing component. This adaptive response demonstrates a practical advantage of joint optimisation over sequential tuning - a sequentially tuned CF component might receive a higher α by default, actively degrading hybrid performance.

The low absolute metric values are consistent with prior work on sparse datasets at this scale. Popularity bias in the test relevance set - where 64% of ratings are 5 stars and the most frequently rated items dominate both training and test signals - compresses the effective difficulty of the task in ways that disproportionately benefit non-personalised baselines on Recall@10.

Future work should address the CF component's poor performance through implicit feedback augmentation, applying SVD++ or Alternating Least Squares on a denser implicit signal derived from review text or browsing events. Additionally, incorporating contextual signals such as item release date may mitigate the temporal mismatch between training interactions from the 1990s and test interactions from 2020–2023.

6. CONCLUSION.

This paper presented a hybrid recommender system combining SVD-based collaborative filtering and TF-IDF-based content-based filtering with joint Bayesian hyperparameter optimisation. The system was evaluated on a large-scale e-commerce dataset exhibiting 99.5% sparsity. Under these conditions, the content-based component achieved consistent personalisation gains over the popularity baseline, while the collaborative filtering component was unable to learn user-specific representations. The Bayesian optimisation framework appropriately down-weighted the non-contributing component, yielding a hybrid that matched pure content-based performance. A three-tier cold-start fallback and a REST API deployment on Hugging Face Spaces demonstrate the system's practical applicability. The modular architecture permits straightforward substitution of individual components as data density increases.

ISSN 3030-3907

REFERENCES

1. G. Linden, B. Smith, and J. York, "Amazon.com recommendations: Item-to-item collaborative filtering," IEEE Internet Computing, vol. 7, no. 1, pp. 76–80, 2003.
2. C. A. Gomez-Uribe and N. Hunt, "The Netflix recommender system: Algorithms, business value, and innovation," ACM Trans. Manag. Inf. Syst., vol. 6, no. 4, pp. 1–19, 2016.
3. Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," IEEE Computer, vol. 42, no. 8, pp. 30–37, 2009.
4. P. Lops, M. de Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in Recommender Systems Handbook, Springer, 2011, pp. 73–105.
5. R. Burke, "Hybrid recommender systems: Survey and experiments," User Modeling and User-Adapted Interaction, vol. 12, no. 4, pp. 331–370, 2002.
6. X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, "Neural collaborative filtering," in Proc. WWW, 2017, pp. 173–182.
7. N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in Proc. EMNLP, 2019, pp. 3982–3992.
8. T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in Proc. KDD, 2019, pp. 2623–2631.
9. J. Hou, Z. Li, H. He, W. Xu, Y. Zhang, W. Wang, and J. McAuley, "Bridging language and items for retrieval and recommendation," arXiv:2403.03952, 2024.
10. A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, "Methods and metrics for cold-start recommendations," in Proc. SIGIR, 2002, pp. 253–260.

ISSN 3030-3907