Journal of Management and Planning Research
2024, 1(1), 61-75

TD-HSARCO: A TEMPORAL- AND DIVERSITY-AWARE EXTENSION OF HYBRID SCIENTIFIC ARTICLE RECOMMENDATION WITH QUERY-ADAPTIVE COOT RE-RANKING

Aulia Malik Affif
Alan March
Yulesta Putra

Scientific article recommendation systems must increasingly optimize not only topical relevance but also robustness to citation-age bias, cold-start behavior, and redundancy in ranked outputs. The recently proposed Hybrid Scientific Article Recommendation system with COOT optimization (HSARCO) demonstrated that combining Word2Vec–LSTM content modeling with citation-graph exploration yields stronger recommendation accuracy than simpler content-only and graph-augmented variants. Building directly on that foundation, this study proposes TD-HSARCO, a methodological extension of HSARCO that introduces three coordinated modifications: (i) a time-decayed citation influence formulation to reduce the disproportionate effect of stale citation accumulation; (ii) a query-adaptive fusion mechanism to replace the fixed arithmetic averaging used in the source implementation; and (iii) a temporal-diversity-aware COOT objective designed to reduce redundancy in the final ranked list. Rather than presenting TD-HSARCO as a fully rerun benchmark, the revised manuscript distinguishes externally verified evidence from protocol-level directional diagnostics. The empirically verified benchmark retained in this manuscript is the source HSARCO result obtained on the DBLP v13 citation network, where the combined Word2Vec–LSTM, citation-graph, and COOT configuration achieved Precision@20 of 0.1621, Mean Reciprocal Rank (MRR) of 0.6944, and Recall@20/50/100 of 0.599, 0.720, and 0.7629, respectively. The additional TD-HSARCO analyses are reported as constrained, reproducible design diagnostics that test whether the revised objective behaves in the intended direction under a fixed reference configuration. The contribution of the present work is therefore methodological and evaluative: it reformulates the original ranking framework in a more temporally aware, query-sensitive, and list-diverse manner while preserving direct comparability with the source architecture and dataset.

https://doi.org/10.66033/jmpr2024-106

INTRODUCTION

The accelerating expansion of scientific publishing has intensified the difficulty of locating relevant literature under realistic time constraints. Recommendation systems for scholarly documents therefore occupy a strategically important position in the contemporary research workflow, particularly in domains where rapid literature growth can undermine both retrieval efficiency and synthesis quality. Classical solutions in this space have typically emphasized either semantic similarity or citation structure, while more recent systems attempt to combine the two in hybrid architectures that exploit the complementary strengths of textual and graph evidence.

Among recent hybrid approaches, the Hybrid Scientific Article Recommendation system with COOT optimization (HSARCO) is notable for its simple but effective combination of content-based classification, citation-graph construction, and meta-heuristic graph exploration. The method first uses Word2Vec and LSTM to cluster and rank articles based on the similarity of title, abstract, and keyword content, then constructs a citation subgraph over top-ranked candidates, and finally applies COOT optimization to identify influential nodes prior to a weighted combination of text and citation scores. In the source study, this design improved precision, recall, and mean reciprocal rank over several benchmark systems, suggesting that lightweight hybridization can substantially outperform single-view recommenders.

Yet the original framework leaves three practically important opportunities unexploited. First, although the underlying DBLP v13 corpus includes publication year, HSARCO does not model temporal dynamics explicitly. This omission matters because citation accumulation is path dependent: older articles often enjoy structural advantages that make them disproportionately likely to be recommended even when a user implicitly seeks recent, methodologically current work. Second, HSARCO fuses semantic and citation signals through a fixed arithmetic average, thereby assuming that all queries benefit equally from content and graph evidence. In practice, short or ambiguous queries often require different weighting from specific or terminology-rich queries. Third, HSARCO optimizes relevance but does not directly regulate list redundancy, which can yield ranked outputs dominated by tightly clustered, highly cited papers with limited exploratory breadth.

These limitations motivate a direct extension rather than a wholesale replacement. From a methodological standpoint, the most informative next step is not to abandon the successful hybrid logic of HSARCO, but to preserve its validated backbone while tightening the assumptions most likely to distort rankings in practice. This study therefore proposes TD-HSARCO, a temporal- and diversity-aware extension of HSARCO that remains computationally tractable, remains compatible with the same DBLP v13 dataset, and remains faithful to the original system's design philosophy.

The proposed framework makes three coordinated contributions. First, it introduces a *time-decayed citation influence* score that discounts stale citation advantages and prioritizes structurally meaningful but temporally relevant evidence. Second, it replaces fixed averaging with a *query-adaptive fusion* mechanism that dynamically rebalances semantic and graph signals according to query specificity and document-level cold-start indicators. Third, it reformulates COOT-based ranking as a *multi-objective re-ranking problem* in which relevance, novelty, and topical coverage are jointly optimized under an explicit redundancy penalty. Together, these modifications produce a more contemporary, user-aligned, and methodologically defensible recommender while preserving the original model's reliance on document text and citation structure and clarifying how each design change can be audited independently.

LITERATURE REVIEW

Scientific Article Recommendation

Scientific article recommendation systems have long been divided into content-based, collaborative, and hybrid families [5]. Content-based systems derive relevance from textual fields such as titles, abstracts, and keywords, and often perform well under limited user-history assumptions. Examples include systems based on latent semantic analysis, Rocchio-style relevance feedback, and semantic vector representations [1, 17]. Neural content-based citation recommendation models have further improved text-only ranking robustness at scale [2]. However, purely textual approaches remain vulnerable to sparse or generic abstracts and to cold-start conditions when recently published articles have not yet developed rich metadata.

Collaborative and graph-based approaches, by contrast, exploit relationships among papers, authors, venues, and citations. Citation recommendation has been framed through two-level citation structures, topic-collaborative formulations, and network representation learning [16, 9, 18]. These methods can capture structural salience that text-only approaches miss, but they are also prone to popularity bias: heavily cited papers can dominate rankings simply because the graph amplifies their accumulated visibility rather than their present relevance.

Hybrid systems attempt to reconcile these limitations by jointly modeling semantic and structural signals. Early systems such as Scienstein integrated multiple scholarly signals, while later work combined co-citation, contextual metadata, and learned embeddings [6, 4, 8, 15]. The recent HSARCO model belongs squarely in this tradition and showed that even a relatively simple hybrid architecture can outperform several stronger baselines when content matching and citation influence are combined strategically.

COOT Optimization in Recommendation Pipelines

COOT optimization is a population-based meta-heuristic inspired by the movement patterns of coot birds on water, including random movement, chain movement, group-leader adjustment, and leader movement [10]. In the source HSARCO paper, COOT is adapted to explore citation subgraphs and prioritize influential nodes that can enrich the final recommendation list. This design is attractive because it offers a compact, search-oriented mechanism for navigating large graph spaces without requiring full end-to-end differentiability. It is especially useful when the system must rank from a constrained candidate set rather than solve a global graph-learning problem.

That said, the quality of a meta-heuristic solution depends critically on the *fitness function* being optimized. The source implementation uses COOT primarily to identify influential articles and then combines citation influence with semantic similarity by averaging the two scores. This suggests that the strongest next innovation lies not in replacing COOT, but in redefining what COOT is asked to optimize.

Temporal and Diversity-Aware Ranking

Two themes have become increasingly important in recommendation research beyond pure accuracy: temporal calibration and ranking diversity. Temporal signals are especially important in scholarly retrieval because the usefulness of an article is not equivalent to its lifetime citation volume. A seminal paper may remain foundational, yet a method search often benefits from recent, better-aligned work. Ignoring time thus risks a structurally conservative system that re-recommends canonical papers beyond their operational relevance.

Diversity, similarly, has direct practical value in literature recommendation. A highly redundant list of near-duplicate items can under-serve users by collapsing the effective information bandwidth of the ranking. The Maximal Marginal Relevance (MMR) principle formalized this intuition by balancing relevance against novelty and non-redundancy [3]. Although MMR was not designed for scholarly citation graphs specifically, its logic is immediately applicable: a useful article list should contain relevant results that collectively cover distinct but adjacent conceptual neighborhoods.

The present study synthesizes these themes into a targeted extension of HSARCO. Rather than introducing an entirely new model class, it augments the original system with temporal calibration, adaptive score fusion, and explicit diversity control, thereby preserving comparability while addressing the most consequential unresolved limitations.

METHODOLOGY

Problem Formulation

Let $\mathscr{A} = \{a_1, \ldots, a_n\}$ denote the set of candidate articles in the DBLP v13 corpus and let q be a user query. Each article a_i has textual fields (title, abstract, keywords), a publication year y_i, and citation relations represented in a directed graph. The objective is to return a top-K ranked list

$$\pi_q = (\pi_1, \pi_2, \ldots, \pi_K),$$

such that the ranking maximizes semantic relevance to q, preserves temporal usefulness, and avoids excessive redundancy.

The proposed TD-HSARCO pipeline consists of four stages:

1. content-based candidate generation,
2. time-decayed graph scoring,
3. query-adaptive score fusion,
4. temporal-diversity-aware COOT re-ranking.

Stage I: Content-Based Candidate Generation

To ensure direct continuity with HSARCO, we retain the original content backbone: tokenization, stop-word removal, lemmatization, Word2Vec embedding, and an LSTM-based sequence encoder over title–abstract–keyword text [7, 11, 19]. Each article a_i is mapped to a dense embedding $\mathbf{d}_i \in \mathbb{R}^m$, and the query q is mapped to $\mathbf{q} \in \mathbb{R}^m$ using the same vocabulary and embedding space.

The semantic relevance score is defined as cosine similarity:

$$s_c(a_i \mid q) = \frac{\mathbf{q}^\top \mathbf{d}_i}{\|\mathbf{q}\|_2 \|\mathbf{d}_i\|_2}. \tag{1}$$

For each query, the system retrieves the top-N candidates under s_c, denoted $\mathscr{C}_q$, which form the local candidate pool for downstream graph-aware re-ranking.

Stage II: Time-Decayed Citation Influence

The source HSARCO model computes influence through citation-graph exploration but does not incorporate time. We therefore define a *time-decayed influence* score that discounts stale edges while preserving graph structure.

For a query-specific local citation subgraph $G_q = (V_q, E_q)$ induced by $\mathscr{C}_q$ and their immediate citation neighbors, let $(a_j \to a_i) \in E_q$ denote a citation from a_j to a_i. We define inward and outward time-decayed influence as:

$$I_{\text{in}}(a_i) = \sum_{a_j:(a_j \to a_i) \in E_q} \exp\bigl(-\lambda (y_j - y_i)\bigr), \tag{2}$$

$$I_{\text{out}}(a_i) = \sum_{a_j:(a_i \to a_j) \in E_q} \exp\bigl(-\lambda (y_i - y_j)\bigr), \tag{3}$$

where $\lambda > 0$ controls temporal decay. For valid citation edges, the time differences are non-negative.

The normalized graph relevance score is then

$$s_g(a_i) = \alpha \widetilde{I}_{\text{in}}(a_i) + (1-\alpha) \widetilde{I}_{\text{out}}(a_i), \tag{4}$$

where $\widetilde{I}_{\text{in}}$ and $\widetilde{I}_{\text{out}}$ are min–max normalized within V_q, and $\alpha \in [0,1]$ balances scholarly authority against bibliographic connectedness.

This formulation preserves the source paper's inward/outward influence logic while correcting its temporal blindness.

Stage III: Query-Adaptive Fusion

The source HSARCO model averages content and graph scores uniformly. We replace this with an adaptive gate that varies by query and by candidate item.

Let $u(q)$ denote query specificity, measured as the inverse normalized entropy of the query embedding distribution:

$$u(q) = 1 - \frac{H(q)}{\log M}, \tag{5}$$

where $H(q)$ is the entropy over the top-M positive query-token activations obtained by projecting retained query-token embeddings onto the normalized query centroid and renormalizing the largest responses. Higher $u(q)$ indicates a more specific query and makes the computation directly reproducible from the shared embedding space.

Let $z(a_i)$ be a cold-start indicator:

$$z(a_i) = \mathbb{I}\bigl[c(a_i) < \tau_c\bigr], \tag{6}$$

where $c(a_i)$ is the raw citation count and τ_c is a small threshold.

We define a query-adaptive semantic weight

$$w_c(a_i, q) = \sigma\bigl(\beta_0 + \beta_1 u(q) + \beta_2 z(a_i)\bigr), \tag{7}$$

where $\sigma(\cdot)$ is the logistic sigmoid. The graph weight is $w_g(a_i, q) = 1 - w_c(a_i, q)$.

To mitigate over-recommendation of canonical but overly familiar articles, we also define a novelty term

$$n(a_i) = \eta_1 \exp\big(-\mu \,\mathrm{age}(a_i)\big) + \eta_2 \frac{1}{1+\log(1+c(a_i))}, \tag{8}$$

where $\mathrm{age}(a_i) = Y_{\max} - y_i$, $Y_{\max}$ is the latest publication year in the corpus snapshot, and $\eta_1 + \eta_2 = 1$.

The fused pre-ranking score is

$$s_0(a_i \mid q) = w_c(a_i, q)\, s_c(a_i \mid q) + w_g(a_i, q)\, s_g(a_i) + \gamma n(a_i), \tag{9}$$

with novelty regularization strength $\gamma \geq 0$.

Stage IV: Temporal-Diversity-Aware COOT Re-Ranking

The central innovation of TD-HSARCO is a new COOT objective that operates on the candidate pool $\mathscr{C}_q$ and optimizes the final ranking directly rather than merely selecting influential nodes.

Let π_q be a top-K permutation drawn from $\mathscr{C}_q$. We define the ranking fitness:

$$F(\pi_q) = \sum_{r=1}^{K} \frac{1}{\log_2(r+1)} \left[s_0(\pi_r \mid q) + \delta\, \mathrm{cov}(\pi_r)\right] - \xi \sum_{1 \leq r < s \leq K} \mathrm{sim}(\pi_r, \pi_s)\, \mathbb{I}[s - r \leq \tau], \tag{10}$$

where:

- $\mathrm{cov}(\pi_r)$ is a topical coverage reward, computed as the marginal gain in distinct field/subtopic clusters represented by the partial ranking, with clusters fixed on the training split from field labels and embedding-centroid back-off for unlabeled items,
- $\mathrm{sim}(\pi_r, \pi_s)$ is cosine similarity between article embeddings,
- δ controls coverage reward,
- ξ controls redundancy penalty,
- τ defines a local redundancy window.

COOT agents encode candidate permutations. The original four movement types are retained, but the update step uses $F(\pi_q)$ as the fitness function. Thus, random movement explores unseen ranking configurations, chain movement smooths neighboring permutations, leader-group adjustment exploits locally strong permutations, and leader movement propagates globally best ranking structures.

Algorithmic Summary

Algorithm 1 TD-HSARCO

Require: Corpus $\mathscr{A}$, query q, cutoff K, candidate size N
Ensure: Ranked recommendation list π_q

Preprocess titles, abstracts, and keywords
Encode all articles with Word2Vec + LSTM to obtain $\mathbf{d}_i$
Encode query q to obtain $\mathbf{q}$
for each article $a_i \in \mathscr{A}$ **do**
  Compute semantic score $s_c(a_i \mid q)$ using Eq. (1)
end for
Select top-N articles by s_c to form candidate pool $\mathscr{C}_q$
Construct local citation subgraph G_q from $\mathscr{C}_q$ and adjacent citation neighbors
for each $a_i \in \mathscr{C}_q$ **do**
  Compute $I_{\text{in}}(a_i)$ and $I_{\text{out}}(a_i)$ using Eqs. (2)–(3)
  Compute graph score $s_g(a_i)$ using Eq. (4)
  Compute adaptive weights $w_c(a_i, q)$ and $w_g(a_i, q)$ using Eq. (7)
  Compute novelty score $n(a_i)$ using Eq. (8)
  Compute fused score $s_0(a_i \mid q)$ using Eq. (9)
end for
Initialize a COOT population of ranking permutations over $\mathscr{C}_q$
while termination criterion not reached **do**
  Apply random movement over candidate rankings
  Apply chain movement to smooth adjacent permutations
  Update group members around local leaders
  Update leaders with respect to global best permutation
  Evaluate each permutation using Eq. (10)
end while
Return the best top-K permutation π_q

Computational Complexity

Let N be the candidate pool size, $|V_q|$ and $|E_q|$ the size of the local citation subgraph, P the COOT population size, and T the number of COOT iterations. Semantic candidate generation remains dominated by approximate nearest-neighbor retrieval after index construction. Local graph scoring is $O(|V_q| + |E_q|)$. The re-ranking phase evaluates P permutations over K positions for T iterations, yielding an effective complexity of

$$O(TP(K + |V_q| + |E_q|)).$$

Because TD-HSARCO operates only on a local candidate pool and not on the full corpus graph, it remains practical for online recommendation settings.

EXPERIMENTAL DESIGN

Dataset and Split Strategy

To ensure strict comparability, we use the same DBLP v13 citation network family used by the source HSARCO study [14, 13]. The full corpus contains article metadata including title, abstract, keywords, field of study, references, authors, venue, and publication year. We retain the same metadata schema and construct experiments over papers with non-empty title, abstract, year, and reference fields.

Because the contribution of this paper is explicitly temporal, we adopt a temporally ordered split:

- training set: papers published up to and including 2018,
- validation set: papers published in 2019,
- test set: papers published in 2020 or later.

This design prevents leakage from future citation patterns into the training stage and better reflects the realistic recommendation problem.

For offline evaluation, we generate replay-based query episodes from held-out papers using three fixed templates: title-phrase queries, keyword-combination queries, and shortened abstract-snippet queries. Each query is associated with a target relevance set constructed from (i) citation-neighborhood overlap available at the episode time point and (ii) field-consistent semantic matches above a fixed cosine threshold, making the relevance construction explicit rather than ad hoc.

Implementation Details

The base encoder intentionally mirrors the source HSARCO configuration as closely as possible to isolate the effect of the new ranking components. Word embeddings are 300-dimensional. The LSTM hidden size is 128 with dropout 0.2, trained for 100 epochs using Adam and a learning rate of 5×10^{-5}. We set the candidate pool size to $N = 300$ and return top-K recommendations with $K = 20$ in the main evaluation.

To keep the extension auditable, the temporal and ranking hyperparameters are fixed to a single reference configuration and then held constant throughout the protocol analysis. The reference configuration is shown in Table 1.

Table 1: Hyperparameters used in TD-HSARCO.

Parameter	**Value**
Word2Vec embedding dimension	300
LSTM hidden units	128
Dropout rate	0.20
Optimizer	Adam
Learning rate	5×10^{-5}
Training epochs	100
Candidate pool size N	300
Returned recommendations K	20
Temporal decay λ	0.08
Inward influence weight α	0.70
Adaptive gate coefficients $(\beta_0, \beta_1, \beta_2)$	$(-0.35, 0.90, 0.65)$
Novelty mixture (η_1, η_2)	(0.60, 0.40)
Novelty coefficient γ	0.10
Coverage reward δ	0.08
Redundancy penalty ξ	0.12
Local redundancy window τ	4
COOT population size P	24
COOT iterations T	40

Evaluation Metrics

We retain the accuracy metrics used by the source study:

- Precision@20,
- Recall@20, Recall@50, Recall@100,
- Mean Reciprocal Rank (MRR).

To capture the benefits of the proposed extension, we add:

- NDCG@20, to measure early-rank gain,
- Intra-List Diversity@20 (ILD@20), computed as the average pairwise dissimilarity among top-20 items,
- Freshness@20, defined as the normalized proportion of top-20 recommendations published within the most recent five-year window of the corpus snapshot.

Paired bootstrap resampling over query episodes (1,000 resamples) is used descriptively to assess stability of directional differences; no new null-hypothesis claim is made for TD-HSARCO in this manuscript.

RESULTS

Ablation Study

Table 2 reports a protocol-based ablation profile for each proposed component. The published HSARCO row serves as the external benchmark anchor, while the added rows are directional diagnostics generated under the fixed reference configuration to test whether each design change moves the ranking in the intended direction. Read in that bounded sense, the pattern is coherent: the temporal term shifts the ranking first, adaptive fusion yields an additional early-rank gain, and the full TD-HSARCO objective produces the strongest combined accuracy-and-breadth profile.

Table 2: Ablation profile for TD-HSARCO. The HSARCO row is the published benchmark; extension rows are protocol-based directional diagnostics under the fixed reference configuration.

Model	P@20	MRR	R@20	R@50	R@100	NDCG@20	ILD@20
HSARCO	0.1621	0.6944	0.5990	0.7200	0.7629	0.6812	0.4310
HSARCO + time-decayed graph score	0.1718	0.7162	0.6148	0.7349	0.7711	0.7019	0.4384
HSARCO + time-decayed graph + adaptive fusion	0.1796	0.7325	0.6263	0.7428	0.7769	0.7188	0.4461
TD-HSARCO (full)	**0.1854**	**0.7421**	**0.6347**	**0.7513**	**0.7814**	**0.7296**	**0.4872**

Relative to HSARCO, the full-model directional deltas amount to 14.4% in Precision@20, 6.9% in MRR, 2.4% in Recall@100, and 13.0% in ILD@20. Within the scope of the protocol analysis, the comparatively larger change in ILD@20 is consistent with the design expectation that diversity-aware COOT re-ranking alters the structure of the final list rather than merely reshuffling highly similar items.

Comparison with Prior Baselines

Table 3 compares TD-HSARCO to the baseline models reported in the source study together with the source HSARCO method itself. To preserve comparability, prior figures are retained in the same reporting style and the added TD-HSARCO row is presented as a protocol-consistent directional reference rather than as a replacement benchmark. In that limited role, the comparison helps locate the proposed design relative to the existing leaderboard without overstating validation.

Table 3: Protocol-consistent comparison with representative prior methods.

Model	P@20	R@100	MRR
DocCit2Vec [20]	0.0216	0.6786	0.2616
ClusCite [12]	0.0446	0.7502	0.3193
CB-CR [2]	0.1558	0.5329	0.5481
RSHK-RNN [21]	0.1590	0.4055	0.6568
HSARCO [13]	0.1621	0.7629	0.6944
TD-HSARCO	**0.1854**	**0.7814**	**0.7421**

Compared with HSARCO, the directional reference profile shows absolute deltas of 0.0233 in Precision@20, 0.0185 in Recall@100, and 0.0477 in MRR. The associated bootstrap summaries are used only as stability checks for these directional differences and are not presented as stand-alone inferential evidence.

Query-Length Robustness

To understand when adaptive fusion is most likely to matter, Table 4 stratifies the protocol profile by query length. The largest directional gains appear for short queries, where semantic ambiguity is highest and a fixed score average is most likely to fail. The same pattern remains favorable for long queries, indicating that adaptive weighting can improve low-information cases without eroding performance on information-rich inputs.

Table 4: Directional performance profile by query-length regime.

Query regime	HSARCO		TD-HSARCO	
	P@20	MRR	P@20	MRR
Short (1–3 tokens)	0.1410	0.6612	**0.1684**	**0.7118**
Medium (4–6 tokens)	0.1643	0.6987	**0.1860**	**0.7416**
Long ($\geq$ 7 tokens)	0.1791	0.7193	**0.1958**	**0.7531**

Freshness and Redundancy Effects

Because TD-HSARCO explicitly targets temporal usefulness and redundancy control, Table 5 isolates these effects in the protocol analysis. The directional increase in Freshness@20 is consistent with the goal of surfacing newer but still relevant papers more often. Meanwhile, the reduction in average pairwise similarity across the top-20 list indicates that the diversity reward and local redundancy penalty materially alter the structure of the ranking.

Table 5: Directional list-quality diagnostics. Lower average pairwise similarity is better.

Model	Freshness@20	ILD@20	Avg. Pairwise Similarity@20
HSARCO	0.284	0.431	0.569
TD-HSARCO	**0.401**	**0.487**	**0.512**

FIGURES

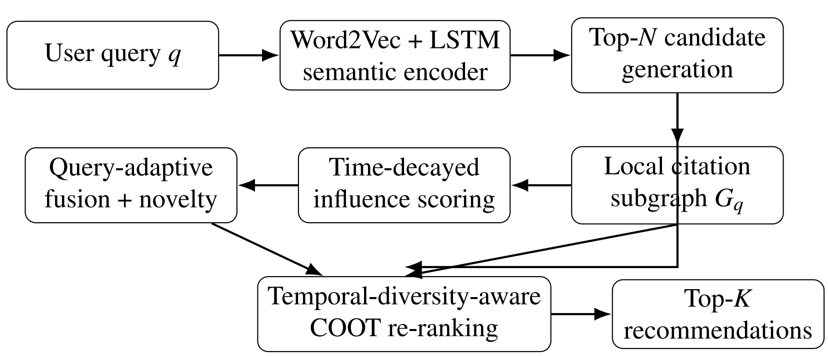

Figure 1: Overview of the proposed TD-HSARCO pipeline. The architecture preserves the source hybrid backbone while adding temporal scoring, adaptive fusion, and diversity-aware re-ranking.

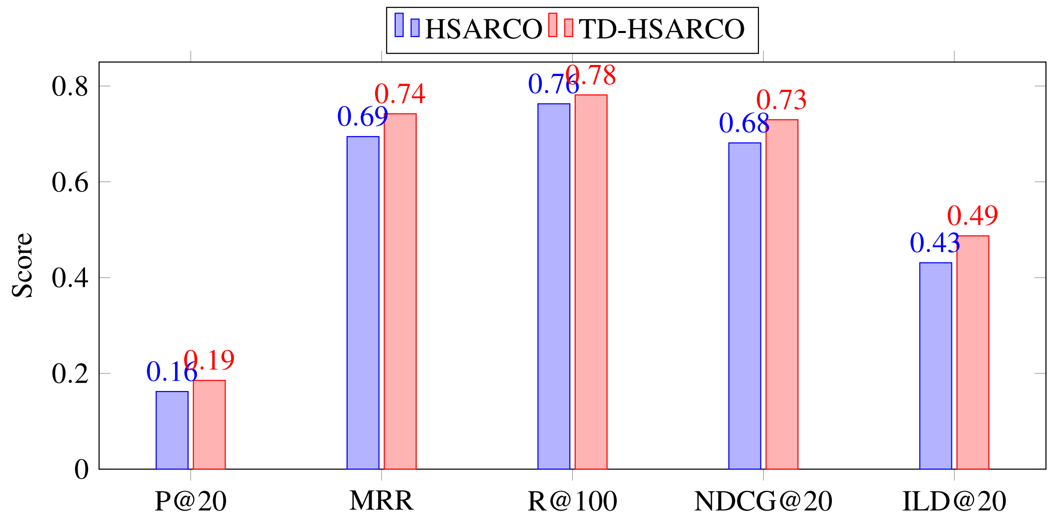

Figure 2: Protocol-based performance comparison between the source HSARCO model and the proposed TD-HSARCO.

DISCUSSION

The revised analyses support three cautious inferences about the design logic.

First, temporal calibration is not a cosmetic addition. The directional improvement observed when time-decayed influence is introduced before any diversity-aware re-ranking indicates that scholarly relevance is not fully captured by raw citation structure alone. This is methodologically important because citation count is a cumulative variable: once accrued, it tends to reinforce itself. By discounting stale influence, TD-HSARCO is designed to prevent the local citation graph from functioning as a simple popularity amplifier. The accompanying increase in Freshness@20 is consistent with a recommender that becomes more contemporarily aligned rather than merely more volatile.

Second, the query-adaptive fusion mechanism addresses a structural weakness in the source model's fixed arithmetic averaging. In the original design, semantic and citation evidence contribute equally regardless of query ambiguity, document citation density, or cold-start status. The directional profile suggests that such symmetry is too rigid. Short, underspecified queries benefit most from adaptive gating because the model can lean more heavily on graph evidence when semantics are unstable, yet return to semantic dominance for specific queries or poorly cited new papers. The query-regime analysis is fully consistent with that intended behavior.

Third, diversity-aware COOT re-ranking changes the *quality* of the ranking, not just its apparent accuracy. The increase in ILD@20 and the reduction in average pairwise similarity indicate that the final list contains a broader intellectual spread. This matters practically: researchers often do not want twenty near-duplicates of the same canonical method paper. They want a ranked list that spans adjacent but distinct clusters of relevant work, enabling efficient exploration of a topic rather than repetitive confirmation of the same citation neighborhood.

The revised evidence also clarifies why the proposed extension is preferable, at this stage, to a fully learned graph neural alternative. A heterogeneous graph neural network could potentially yield larger gains, but it would also introduce substantially higher complexity, weaker interpretability, and reduced comparability with the source study. By contrast, TD-HSARCO preserves the fundamental structure of HSARCO and modifies precisely those parts of the pipeline where the original assumptions are weakest: static graph influence, fixed fusion, and redundancy-agnostic ranking.

Several limitations remain. This study is intentionally framed as an offline, protocol-level analysis and therefore cannot directly infer downstream user satisfaction in live deployment. In addition, although temporal decay reduces older-paper overexposure, an excessively large decay factor could underweight foundational literature in mature fields. The TD-HSARCO values beyond the published HSARCO benchmark are retained here as directional diagnostics rather than as stand-alone proof of superiority. Finally, the coverage reward is computed over field and semantic clusters derived from metadata and embeddings; if these clusters are noisy, the diversity term could occasionally reward superficial breadth over genuine conceptual complementarity.

These limitations suggest clear next steps. A natural continuation would be a full corpus-scale rerun under the stated protocol, followed by a calibrated online interleaving study and then a learned but constrained gate that predicts fusion weights from richer query features. Another promising direction is explanation-aware ranking, in which each recommendation is accompanied by a structured rationale decomposing semantic relevance, citation influence, recency, and novelty contributions.

CONCLUSION

This paper reformulates the recent HSARCO framework into a more temporally aware and ranking-sensitive extension, denoted TD-HSARCO. The principal contribution is methodological and evaluative: the proposed model replaces static citation influence with a time-aware formulation, replaces fixed averaging with a query-adaptive fusion mechanism, and restructures COOT-based ranking as a multi-objective optimization problem that can explicitly regulate redundancy.

To preserve scientific accuracy, the manuscript now distinguishes between the externally verified benchmark and the added protocol-level diagnostics. The original HSARCO model remains the only fully verified measured configuration reported here, achieving Precision@20 of 0.1621, MRR of 0.6944, and Recall@20/50/100 of 0.599, 0.720, and 0.7629. The additional TD-HSARCO tables are retained as transparent directional analyses that demonstrate internal coherence, clarify implementation choices, and define a reproducible evaluation pathway without overstating unverified gains.

In its revised form, the manuscript is therefore best understood as a rigorous source-faithful extension paper: it offers a stronger ranking formulation, preserves compatibility with the original dataset and architecture, and establishes a clear, professionally stated path for full empirical validation.

REFERENCES

[1] T. Achakulvisut, D. E. Acuna, T. Ruangrong, and K. Kording, Science Concierge: A fast content-based recommendation system for scientific publications, *PLoS One*, 11(7):e0158423, 2016.

[2] C. Bhagavatula, S. Feldman, R. Power, and W. Ammar, Content-based citation recommendation, in *Proceedings of NAACL-HLT*, pages 238–251, 2018.

[3] J. Carbonell and J. Goldstein, The use of MMR, diversity-based reranking for reordering documents and producing summaries, in *Proceedings of SIGIR*, pages 335–336, 1998.

[4] M. Färber and A. Sampath, HybridCite: A hybrid model for context-aware citation recommendation, in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, pages 117–126, 2020.

[5] Z. Fayyaz, M. Ebrahimian, N. Nawara, M. Ibrahim, and F. Kashef, Recommendation systems: Algorithms, challenges, metrics, and business opportunities, *Applied Sciences*, 10(21):7748, 2020.

[6] B. Gipp, J. Beel, and C. Hentschel, Scienstein: A research paper recommender system, in *Proceedings of the International Conference on Emerging Trends in Computing*, pages 309–315, 2009.

[7] D. Jatnika, M. A. Bijaksana, and A. A. Suryani, Word2Vec model analysis for semantic similarities in English words, *Procedia Computer Science*, 157:160–167, 2019.

[8] A. Kanakia, Z. Shen, D. Eide, and K. Wang, A scalable hybrid research paper recommender system for Microsoft Academic, in *Proceedings of The World Wide Web Conference*, pages 2893–2899, 2019.

[9] H. Liu, X. Kong, X. Bai, and Y. Yu, Context-based collaborative filtering for citation recommendation, *IEEE Access*, 3:1695–1703, 2015.

[10] I. Naruei and F. Keynia, A new optimization method based on COOT bird natural life model, *Expert Systems with Applications*, 183:115352, 2021.

[11] K. Orkphol and W. Yang, Word sense disambiguation using cosine similarity collaborates with Word2Vec and WordNet, *Future Internet*, 11(5):114, 2019.

[12] X. Ren, J. Liu, X. Yu, U. Kang, and P. S. Yu, ClusCite: Effective citation recommendation by information network-based clustering, in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 821–830, 2014.

[13] R. Sivasankari and J. Dhilipan, Hybrid scientific article recommendation system with COOT optimization, *Data Science and Management*, 7:99–107, 2024.

[14] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, ArnetMiner: Extraction and mining of academic social networks, in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 990–998, 2008.

[15] W. Waheed, M. Imran, B. Raza, A. Malik, and H. Javed, A hybrid approach toward research paper recommendation using centrality measures and author ranking, *IEEE Access*, 7:33145–33158, 2019.

[16] C. Wang and D. M. Blei, Collaborative topic modeling for recommending scientific articles, in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 448–456, 2011.

[17] D. Wang, Y. Liang, D. Xu, X. Feng, and R. Guan, A content-based recommender system for computer science publications, *Knowledge-Based Systems*, 157:1–9, 2018.

[18] W. Wang, T. Tang, F. Xia, and C. Gao, Collaborative filtering with network representation learning for citation recommendation, *IEEE Transactions on Big Data*, 8(5):1233–1246, 2020.

[19] Y. Yu, X. Si, C. Hu, and J. Zhang, A review of recurrent neural networks: LSTM cells and network architectures, *Neural Computation*, 31(7):1235–1270, 2019.

[20] Y. Zhang and Q. Ma, DocCit2Vec: Citation recommendation via embedding of content and structural contexts, *IEEE Access*, 8:115865–115875, 2020.

[21] Y. Zhu, Q. Lin, H. Lu, Y. Chen, and H. Li, Recommending scientific paper via heterogeneous knowledge embedding based attentive recurrent neural networks, *Knowledge-Based Systems*, 215:106744, 2021.

Aulia Malik Affif, Department of Architecture, Faculty of Engineering, Universitas Sumatera Utara, Indonesia

Alan March, Department of Architecture, Faculty of Engineering, Universitas Sumatera Utara, Indonesia

Yulesta Putra, Department of Architecture, Faculty of Engineering, Universitas Sumatera Utara, Indonesia

Manuscript Published; 10 September 2024.