Journal of Scientometric Research, 2026; 15(2):386-397.
https://www.jscires.org

Research Article

A Contrastive Supervised Learning Framework for Reliable Reviewer-Manuscript Matching

Nishith Kotak[1,*], Bakul Gohel[2], Arjav Bavarva[1]

[1]Department of Information and Communication Technology, Marwadi University, Rajkot, Gujarat, INDIA.
[2]Department of CE-AI and Big Data, Marwadi University, Rajkot, Gujarat, INDIA.

ABSTRACT

Automated reviewer-manuscript matching is a core element of contemporary peer-review systems, often posed as a similarity-based retrieval task on pretrained text embeddings. Although the current methods perform well on ranking measures like Mean Reciprocal Rank (MRR) and NDCG, current methods implicitly posit that the closer the similarity, the greater the assignment confidence. In this work, we demonstrate that this assumption is actually flawed because of the role of embedding geometry that is often ignored. With a thorough analysis of various transformer-based encoders and a variety of scientific fields, we show that pretrained embeddings possess strong inter-intra domain overlap, anisotropy, and weak separability, which results in ambiguous similarity distribution and unreliable assignment. We also demonstrate that traditional ranking measures do not reflect confidence in margins and absolute separability, and tend to hide geometric flaws. In order to overcome these constraints, we introduce a Contrastive Geometry Refinement Framework, which explicitly re-structures the embedding space on the basis of supervised contrastive learning. The proposed solution creates intra-domain compactness and inter-domain separation, which leads to better centroid separation, larger similarity margins and higher discriminability. Theoretical studies prove that contrastive optimization enhances inter-domain separation and the intra-domain variance is decreased. Empirical findings show that there are significant improvements in many metrics, such as almost 2 times improvement in MRR and NDCG, amplification in margins, and thereby geometric disentanglements in domains. We find that incorporating geometry and not similarity magnitude is the key predictor of trustworthy reviewer assignment. This paper offers a principled basis to enhance performance and trust in scholarly recommendation systems by redefining reviewer recommendation as a geometry-guided contrastive learning problem.

Keywords: Contrastive Learning, Embedding Geometry, Geometric Separability, Reviewer Matching, Semantic Representations.

Correspondence:
Nishith kotak
Department of Information and Communication Technology, Marwadi University, Rajkot-360003, INDIA.
Email: nishith.kotak@marwadieducation.edu.in

Received: 13-03-2026;
Revised: 08-05-2026;
Accepted: 21-07-2026.

INTRODUCTION

Peer-review process plays a key role in ensuring the quality, rigor and integrity of scientific publication. Nevertheless, the fast increase in the number of manuscripts submitted has resulted in a significant computational problem of reviewer assignment. This issue is solved by modern automated systems which encode manuscripts and reviewer profiles into dense vectors and find the appropriate reviewers via scores of similarity.

Existing approaches used the matching of keywords and hand-crafted expertise modeling (Goldsmith and Sloan, 2007) whereas the recent techniques use transformer-based encoders, such as BERT (Devlin *et al.*, 2019) and domain-specific versions, including SciBERT (Beltagy *et al.*, 2019) BioBERT (Lee *et al.*, 2020), MathBERT (Reusch, 2022), PhysBERT (Hellert *et al.*, 20 24), and FinBERT (Liu *et al.*, 2021). Even better semantic representations are ensured by citation-aware models like SPECTER (Cohan *et al.*, 2020), which use citation signals to enhance semantic representations. These approaches usually define the problem of reviewer assignment as a similarity-based retrieval problem, which is measured by ranking measures, like Mean Reciprocal Rank (MRR) and Normalized Discounted Cumulative Gain (NDCG) (Wang *et al.*, 2023).

Although these metrics are performing well, the current methods implicitly believe that the greater the similarity, the greater the confident of the assignment. In this piece we demonstrate the fact that such an assumption is flawed. Similarity scores are effective in capturing pairwise alignment, but do not provide the overall geometric structure of the embedding space, which is essential to confident discrimination between relevant and competing reviewers.

DOI: 10.5530/jscires.20260164

We find that pretrained embeddings have a large inter-intra domain overlap, anisotropy, and low separability through empirical analysis across a variety of encoders and scientific domains. Consequently, similarity distributions tend to be dense and overlapping, thus, resulting in low margins between high-ranking candidates and indeterminate assignment decisions. Notably, we demonstrate the insensitivity of such metrics as MRR and NDCG to such geometric shortcomings since they do not quantify absolute separability or confidence but relative ranking.

These observations demonstrate a crucial weakness of similarity-based retrieval the ability to rank correctly does not imply accurate assignment. Specifically, in the case of poorly structured embeddings, high similarity scores can be obtained due to overlapping distributions instead of semantic similarity, leading to unstable and low-confidence recommendations.

As a solution to these shortcomings, we introduce a Contrastive Geometry Refinement Framework which explicitly redefines the embedding space with a supervised contrastive learning. The method proposed learns a geometry-aware transformation that imposes a compactness within-domain and dispersion between-domain constraints, and thus enhances the discriminative structure of the representation space.

Our framework is also built upon using margin-based and distribution-aware signals, unlike the traditional approaches which use similarity alone to make assignment decisions, which is more reliable.

We give a theoretical treatment that shows that centroid separation is enhanced and intra-domain variance is minimized under supervised contrastive learning resulting in better geometric separability. Experimental evidence on a variety of domains and encoder designs indicates significant improvements in ranking performance, such as an almost 2 × improvement in MRR and NDCG, and more pronounced amplification of similarity margins and the existence of distinct decision boundaries.

Overall, this work confirms that embedding geometry, and not similarity magnitude alone is the main determinant of trustful reviewer assignment. Reforming the notion of reviewer recommendation into a geometry sensitive contrastive learning task, we offer a principled approach to enhancing performance and confidence in scholarly matching systems.

This paper makes the following key contributions

- We propose a novel Contrastive Geometry Refinement Framework that learns a geometry-aware transformation of embeddings using supervised contrastive learning to enforce domain-level separability.
- We theoretically establish that supervised contrastive learning increases centroid separation while reducing intra-domain variance, providing a principled foundation for improved assignment reliability.
- We demonstrate through extensive experiments that the proposed approach significantly improves ranking performance, margin-based confidence, and embedding separability across multiple scientific domains.

RELATED WORK

The reviewer-manuscript matching has been an object of continuing interest because of its key role in large scale peer-review systems. Current literature covers recommendation strategies, academic representation learning and evaluation approaches, but does not pay much attention to the importance of integrating geometry into assignment assurance.

Initial systems used overlapping keywords, expert profiles that were curated, and heuristics rules (Goldsmith and Sloan, 2007; Mimno and McCallum, 2007). Although interpretable, these methods had problems with vocabulary and semantic ambiguities. Manuscripts and reviewer profiles were subsequently modeled using the content-based methods by employing the use of vector-space representations based on titles, abstracts, and publication histories of manuscripts and reviewers respectively (Tang *et al.*, 2012). Graph-based methods also used citation or co-authorship networks to add contextual matching (Charlin and Zemel, 2013; Long and Wong, 2013), but were limited by the sensitivity of handcrafted features to sparse data.

The recent developments are characterized by neural and transformer-based models, which encode manuscripts and reviewer profiles into dense embeddings to be retrieved based on their similarity to each other (Cohan *et al.*, 2020; Devlin *et al.*, 2019). These methods are scalable to real world systems (Wang *et al.*, 2023), however, naturally presuppose that the greater similarity, the greater the assignment confidence, and assessment is mostly limited to ranking statistics such as MRR@K and NDCG@K.

Representation learning has since evolved through TF-IDF and Latent Semantic Analysis (LSA) (Deerwester *et al.*, 1990; Salton *et al.*, 1975) to contextual embeddings like BERT and scientific versions like SciBERT (Beltagy *et al.*, 2019; Devlin *et al.*, 2019). Domain-specific models BioBERT (Lee *et al.*, 2020), can perform better within their domain, but do not generalize well to other domains (Wang *et al.*, 2023). SPECTER also builds on this paradigm by adding citation-aware pretraining (Cohan *et al.*, 2020), but its performance is measured in terms of retrieval, instead of structural embedding properties.

Recent studies have delved into scholarly recommendation and representation learning by using bibliometric, citation-based, and multi-feature modeling methods. An example is a research paper recommendation system proposed by Kanwal and Amjad (Kanwal and Amjad, 2024), which combines several

signals based on citation networks to enhance the quality of matching, whereas Zhang and Wu (Zhang and Wu, 2024), research multi-model frameworks to capture cross-domain scholarly impact by using early citation dynamics. Likewise, Kullmann and Weimer (Kullmann and Weimer, 2024) create an organized scientometric model of scholarly resource modeling, with an accent on systematic representation of academic entities. Together, these methods show the usefulness of using heterogeneous bibliometric and structural properties to do large-scale scholarly recommendation.

Nevertheless, even with these developments, current approaches are still based on similarity-based formulations, in which similar decisions are motivated by pairwise similarity scores based on textual or network-based features. These methods do not explicitly examine the geometry of the underlying representation space, such as intra-domain compactness, inter-domain separation and distributional overlap. Therefore, high similarity scores can still be associated with ambiguous assignments in case the representations of competing candidates are close to each other. This drawback points to the necessity of geometry-conscious learning paradigms beyond similarity-based evaluation to structurally aware and confidence-conscious decision-making, which is sought in the framework presented in the proposed research.

Although these developments have been made, the current evaluation practices are still similarity-based. Cosine similarity and topK accuracy measures evaluate relative ranking, but do not evaluate assignment reliability, especially when the relative rankings of competing candidates are similar. Importantly, they fail to consider the inclusion of space characteristics including the separation of domains within the inter-domain, the compactness of domains in intra-domain as well as distributional overlap, which directly determine the strength of decisions.

More recent neural methods learn reviewer assignment as an end-to-end learning task, learning to match performances with deep encoders optimally in terms of matching performance measures of the task at hand (Zhang *et al.*, 2021). Large-scale systems and benchmarks are further concerned with scalability and deployment realism (Wang *et al.*, 2023) and hybrid designs integrate domain-specific features and metadata (Gu *et al.*, 2022; Wang *et al.*, 2024). However, these approaches still consider similarity as a proxy to the confidence of assignments.

One of the key limitations throughout the previous literature is that there is no clear geometric analysis of embedding spaces. Current methods do not consider intra-domain compactness, inter-domain separation, and anisotropy, nor do they include distribution-sensitive measures of confidence. Consequently, embeddings can have high similarity scores and yet generate ambiguous assignments as a result of shared similarity distributions.

To resolve such drawbacks, we propose a contrastive learning framework that is geometry-aware (imposing structural constraints in the embedding space). Our method represents a step beyond similarity-based retrieval and reinterprets reviewer assignment as a geometric decision problem by integrating supervised contrastive learning, centroid-based matching, and confidence-aware scoring, which facilitates better separability and increases the reliability of assignment results. Detailed Comparison of existing ork with the proposed geometry-aware contrastive framework is given in Table 1.

Geometry-Aware Contrastive Reviewer Assignment Framework

Problem Setup

Accurate reviewer assignment depends on the quality of the underlying representation space in which manuscripts and reviewers are embedded. Given textual descriptions of manuscripts and reviewer expertise, an effective system requires a shared embedding space where semantic proximity reflects domain relevance and expertise alignment.

Let $\mathscr{M} = \{m_1, \ldots, m_n\}$ denote a set of manuscripts and $\mathscr{R} = \{r_1, \ldots, r_k\}$ denote a pool of candidate reviewers. We define a parametric encoder f_θ: $T \rightarrow \mathrm{R}^d$ that maps textual inputs into a d-dimensional embedding space:

$$\mathbf{z}_m = f_\theta(m), \quad \mathbf{z}_r = f_\theta(r)$$

To ensure scale-invariant similarity comparisons and stable optimization, embeddings are normalized onto a unit hypersphere:

$$\mathbf{z}_m = \frac{\mathbf{z}_m}{\| \mathbf{z}_m \|_2}, \quad \mathbf{z}_r = \frac{\mathbf{z}_r}{\| \mathbf{z}_r \|_2}$$

Similarity between a manuscript and a reviewer is computed using cosine similarity:

$$S(m, r) = \mathbf{z}_m^\top \mathbf{z}_r$$

The reviewer assignment problem is formulated as a maximum inner product search (MIPS):

$$\boldsymbol{r}^*(\boldsymbol{m}) = \arg\max_{\boldsymbol{r} \in \mathcal{R}} \boldsymbol{S}(\boldsymbol{m}, \boldsymbol{r})$$

Although this formulation represents pairwise semantic alignment, it is not sensitive to the relative order of the embedding space. Specifically, similarity scores by themselves do not indicate the extent to which a candidate reviewer is discriminated against competing options, and as such, assignment decisions can be ambiguous.

As shown in Figure 1, the proposed framework involves two main steps: (i) geometry-aware representation learning using contrastive refinement, and (ii) confidence-aware reviewer assignment using centroid-based similarity and ranking. The model, beginning

Table 1: Comparison of Prior Work with the Proposed Geometry-Aware Contrastive Framework.

Sl. No.	Objective	Core Idea	Challenges	References
1.	Content-based reviewer assignment	Surface-level textual similarity	Ignores semantic depth and confidence	(Tang *et al.*, 2012)
2.	Topic-model based matching	Probabilistic topic alignment	Topic overlap ≠ assignment reliability	(Charlin and Zemel, 2013)
3.	Scientific text representation	Domain-adapted pretraining (SciBERT)	Not optimized for cross-domain separability	(Beltagy et al., 20 19)
4.	Citation-aware embeddings	Graph-informed representation learning	Focus on retrieval, not assignment confidence	(Cohan *et al.*, 2020)
5.	Neural reviewer matching	End-to-end deep similarity modeling	Treats similarity as confidence proxy	(Zhang *et al.*, 2021)
6.	Benchmark datasets	Large-scale realistic evaluation	Ranking-focused, lacks geometric analysis	(Stelmakh *et al.*, 2023)
7.	Industrial recommender systems	Scalable matching pipelines	No embedding geometry characterization	(Wang *et al.*, 2023)
8.	Domain-specific encoders	Specialized pretrained models	Over-specialization reduces generalization	(Gu *et al.*, 2022)
9.	Embedding benchmarking	Task-based evaluation frameworks	Ignores distributional spread and anisotropy	(Singh *et al.*, 2023)
10.	Hybrid matching systems	Text + metadata fusion	Confidence remains implicit and uncalibrated	(Wang *et al.*, 2024)
Ours	Geometry-aware reviewer assignment	Contrastive refinement + centroid-based scoring	Scarcity of high-quality supervision labels	-

with pretrained embeddings, learns a projection that transforms the embedding space to enhance domain-level separability. Similarity scores, margins and distribution-aware confidence scores are then computed using the refined embeddings, which together decide the final reviewer assignment. The projection network and the dual-objective optimization, together impose intra-domain compactness and inter-domain separation.

Assignment Confidence via Margin and Density

To ensure a reliable reviewer assignment, it is necessary that the similarity is high and that the competition between the candidates is well distinguished. Thus, it is necessary to include local ranking structure and global similarity distribution in assignment confidence.

Margin-Based Confidence

From a ranking perspective (Burges *et al.*, 2005), robust assignment requires a significant gap between the top-ranked candidate and its nearest competitor. We define the margin as:

$$\Delta(m)=S(m,r_{(1)})-S(m,r_{(2)})$$

Where $r_{(1)}$ and $r_{(2)}$ denote the top-1 and top-2 ranked reviewers, respectively. Larger margins indicate more confident assignments, whereas small margins reflect ambiguity due to overlapping similarity scores.

Density-Normalized Confidence

With dense similarities distributions or highly overlapping distributions, margin is inadequate. In order to consider this, we normalize similarity scores with the help of distributional statistics (Hogg *et al.*, 2013).

The mean and variance of similarity scores are defined as:

$$\mu_m = \frac{1}{K}\sum_{r\in\mathcal{R}} S(m,r)$$

$$\sigma_m^2 = \frac{1}{K}\sum_{r\in\mathcal{R}} (S(m,r)-\mu_m)^2$$

We define standardized confidence as:

$$C(m) = \frac{S(m,r_{(1)})-\mu_m}{\sigma_m}$$

This formulation quantifies the extent of the deviation of the top-ranked reviewer with the overall similarity distribution, and it quantifies the relative ranking as well as global separability.

Geometric Structure of Embedding Space

The geometry of the embedding space is the key determinant of the similarity-based retrieval effectiveness. In inter-domain similarity, strong intra-domain compactness and low inter-domain similarity would be needed to guarantee the formation of relevant reviewers in coherent clusters as per (Jain, 2010) that are well separated across other domains.

Domain Separation

We define intra-domain compactness for domain d as:

$$\mathcal{C}_{\text{intra}}^{(d)} = \frac{1}{|\mathcal{R}^{(d)}|^2} \sum_{i,j \in \mathcal{R}^{(d)}} \mathbf{z}_{r_i}^{\top} \mathbf{z}_{r_j}$$

Inter-domain similarity between domains d_1 and d_2 is defined as:

$$\mathcal{C}_{\text{inter}}^{(d_1,d_2)} = \frac{1}{|\mathcal{R}^{(d_1)}||\mathcal{R}^{(d_2)}|} \sum_{i,j} \mathbf{z}_{r_i}^{\top} \mathbf{z}_{r_j}$$

A necessary condition for reliable assignment is:

$$\min_{d} \mathcal{C}_{\text{intra}}^{(d)} > \max_{d_1 \neq d_2} \mathcal{C}_{\text{inter}}^{(d_1,d_2)}$$

Violation of this condition leads to overlapping similarity distributions and ambiguous assignments.

Spectral Geometry

In addition to the pairwise similarity, the covariance structure defines the quality of global embedding. Pretrained embeddings tend to be anisotropic (Ethayarajh, 2019), where the variance is concentrated on a limited set of strong directions.

We define the covariance matrix:

$$\Sigma_R = \frac{1}{K} \sum_{r \in \mathcal{R}} (\mathbf{z}_r - \bar{\mathbf{z}}_R)(\mathbf{z}_r - \bar{\mathbf{z}}_R)^{\top}$$

$$\bar{\mathbf{z}}_R = \frac{1}{K} \sum_{r \in \mathcal{R}} \mathbf{z}_r$$

Anisotropy is quantified using the spectral condition number (Jolliffe, 2011):

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}$$

Lower κ indicates a more isotropic and discriminative embedding space, which is desirable for reliable retrieval.

Contrastive Geometry Refinement

Untrained embeddings do not explicitly separate domains, resulting in overlapping and low-resolution representation spaces. To overcome this issue, we propose a learnable transformation that transforms the embedding geometry following (Chen *et al.*, 2020; Khosla *et al.*, 2020) based on supervised contrastive learning.

Projection Network

The projection network maps embeddings into a geometry-refined latent space:

$$h_1 = xW_1 + b_1$$
$$h_2 = \text{BatchNorm}(h_1)$$
$$h_3 = \text{ReLU}(h_2)$$
$$h_4 = \text{Dropout}(h_3)$$
$$h_5 = h_4 W_2 + b_2$$

The output is normalized:

$$z_i = \frac{h_5}{\| h_5 \|_2}$$

Supervised Contrastive Objective

We adopt supervised contrastive learning to explicitly enforce intra-domain compactness and inter-domain separation (Khosla *et al.*, 2020).

$$\mathcal{L}_{\text{SupCon}} = \sum_{i=1}^{B} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p/\tau)}{\sum_{a \neq i} \exp(z_i \cdot z_a/\tau)}$$

This objective pulls semantically similar instances closer while pushing dissimilar instances apart, directly improving geometric separability (Wang and Isola, 2020).

Centroid Separation under Contrastive Learning

Let the centroid of dom $\mu_c = \frac{1}{|\mathcal{D}_c|} \sum_{i \in \mathcal{D}_c} z_i$

We define centroid separation:

$$\mathcal{S}(c_1, c_2) = \| \mu_{c_1} - \mu_{c_2} \|_2^2$$

Theorem 1 (Centroid Separation under SupCon). Minimizing $\mathcal{L}_{\text{SupCon}}$ increases expected centroid separation between domains while reducing intra-domain variance.

Supervised contrastive learning maximizes similarity among positive pairs and minimizes similarity among negatives (Khosla *et al.*, 2020). Consequently:

$$\mathbb{E}_{i \in c_1, j \in c_2} [z_i^{\top} z_j] \downarrow$$

which reduces cross-domain similarity. Since centroids are averages, this increases $\|\mu_{c1} - \mu_{c2}\|^2$. Simultaneously, intra-domain contraction reduces variance.

This proposed approach assumes the availability of domain-level supervision for contrastive learning. In practice, these labels can be obtained from various sources such as subject categories associated with authors, conference tracks, journal classifications, or metadata-driven signals like keywords or citation networks.

On the other hand, real-world label annotation tends to be noisy, overlapping, or coarse grained. Our Supervised contrastive objective only consider pairwise relationships between samples at the mini-batch level, leading to less sensitivity to global label noise. Moreover, the distillation process builds on top of pretrained embeddings already captures potential semantic information, thus reducing the reliability over supervision. When labels are coarse or partially noisy, the model continues to improve representation quality, although with reduced gains.

Geometry-Aware Assignment Function

We define a unified scoring function that integrates similarity, margin, and distribution-aware confidence:

$$\Phi(m,r) = \alpha S(m,r) + \beta \Delta(m) + \gamma C(m) + \delta G(Z_R)$$

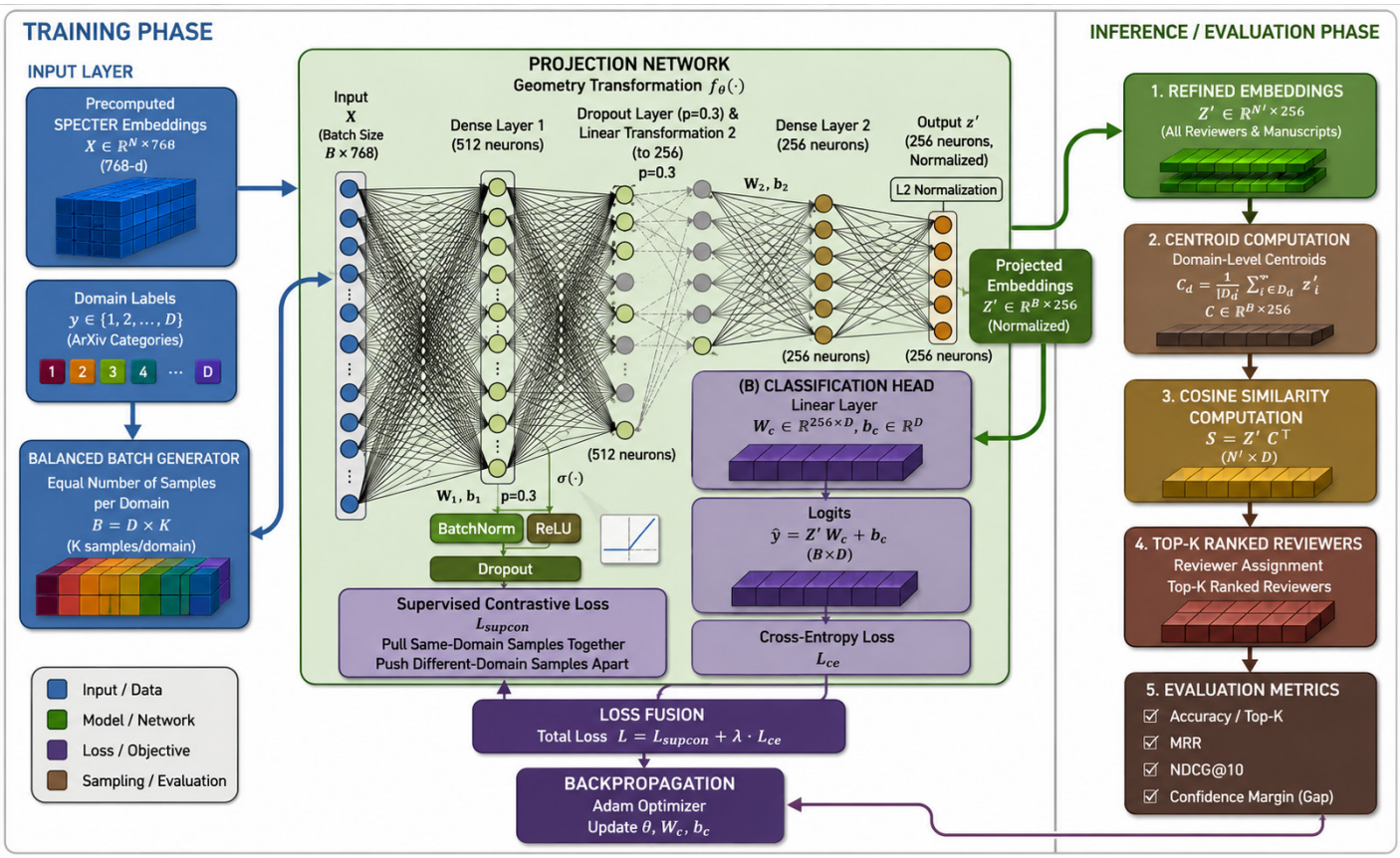


Figure 1: Proposed Contrastive Geometry Refinement framework for reviewer-manuscript matching. The model refines pretrained embeddings through a projection network trained with supervised contrastive and classification objectives, followed by centroid-based similarity computation and confidence-aware ranking.

Where $G(Z_R)$ captures global geometric properties of the embedding space.

The proposed formulation does not enforce hard-domain assignments, based on the nearest mapping cluster centroid. Instead, domain centroids act as geometric reference anchors for structuring the embedding space. As a result, inter-disciplinary manuscripts that span multiple domains are represented in intermediate regions of the embedding space. This design ensures that interdisciplinary manuscripts are not constrained by rigid domain boundaries.

The final assignment is:

$$r^{\dagger}(m) = \text{argmax}\Phi(\text{m,r})\ r \in R$$

This formulation ensures that assignment decisions are not solely based on pairwise similarity, but also incorporate margin-based confidence and global geometric structure, leading to more reliable and discriminative reviewer selection.

RESULTS AND ANALYSIS

Similarity Distributions and Domain Overlap

Our analysis starts with the intrinsic geometric structure of different pretrained embeddings on various domains. The Geometric Separability Condition is violated majorly in accordance with empirical observation:

$$\min_d \mathcal{C}_{\text{intra}}^{(d)} \not> \max_{d_1 \neq d_2} \mathcal{C}_{\text{inter}}^{(d_1,d_2)}$$

Given general-purpose (BERT) and domain-specialized encoders (BioBERT, MathBERT, PhysBERT, FinBERT), the inter-intra similarity gaps are critically small, i.e.<0.02, showing a great deal of overlap between intra-domain and inter-domain similarity distributions.

Table 2 shows the statistics of model-wise cosine similarity and ranking performance. It is possible to note that the average cosine similarity values are low in all domains, and a few of them have near-zero or negative similarity (e.g., SPECTER-EESS, MathBERT), which can be attributed to the lack of semantic cohesion and the instability of embedding alignment.

In extreme cases, MathBERT exhibits negative separation:

$$\mathcal{C}_{\text{intra}}^{\text{Math}} < \mathcal{C}_{\text{inter}}^{\text{CS-Math}}$$

Demonstrating complete geometric ambiguity.

The highest separability of pretrained models is observed in SPECTER in terms of higher values of cosine similarity and ranking performance (MRR@10=0.3986, NDCG@10=0.4118). Its performance, however, differs considerably in terms of the

fields, especially deteriorates in mathematics and EESS, which does not point to consistent geometric structure.

Distributional differences between intra- and inter-domain similarities are statistically proven by statistical tests (Mann-Whitney U, Kolmogorov-Smirnov; $p<0.001$). Nevertheless, the magnitude of the effect is quite low (Cohen's $d\approx0.27$) meaning that there is a significant overlap in practice.

In terms of decision making, this overlap creates small similarity margins among the top-ranked candidates, which makes it have ambiguous assignments although there might be reasonable ranking performance. These results point to a failure of pretrained embeddings to arrange the representation space in a geometrically discriminative way.

Ranking Metrics versus Geometric Separation

We consider the effect of geometry embedding on ranking-based retrieval in terms of MRR10 and NDCG10. Table 2 indicates that models that are comparatively well-structured geometric-wise perform well on the ranking task.

SPECTER has the highest performance of the pretrained models, whereas MathBERT is the worst-performing model since the similarity distributions are not well-separable and unstable. Correlation analysis reveals that there is a strong positive association between geometric separation and ranking measures (Spearman $\rho\approx0.74$ with MRR and $\rho\approx0.77$ for NDCG).

Nevertheless, this relationship is not enough to assign reliably. Ranking measurements are based on relative order, but do not provide absolute detachability and confidence of decision. As noted in Table 2, even the models that have competitive MRR and NDCG scores have low values of absolute similarity and large domain variance.

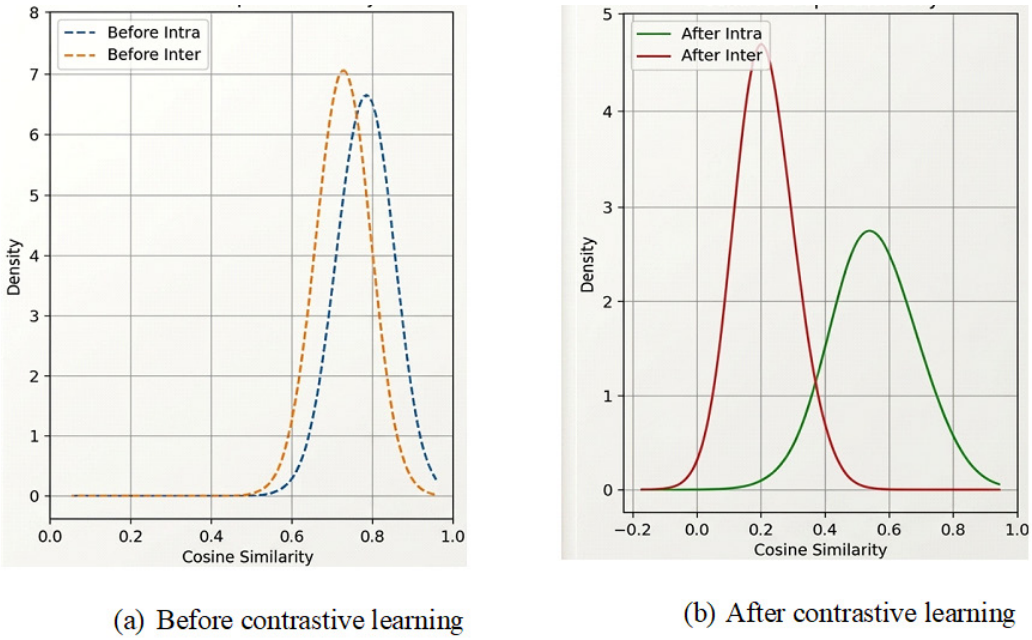
This implies that right reviewers are not highly differentiated with competing candidates and hence small margins and indeterminate assignments. Thus, prioritizing the performance will not ensure credible reviewer choice.

Domain Specialization and Cross-Domain Robustness

The domain-specific encoders are more compact with respect to intra-domain because they are exposed to domain-specific corpora. They however do not extrapolate with heterogeneous reviewer pools in which cross-domain similarity is an important factor.

Although in-domain retrieval performance is improved moderately (~3.6%-4.4%), cross-domain interference still exists because similarity distributures overlap across domains. This contributes to a situation in which reviewers in related but different areas would score high similarity, and assignment reliability will be low.

The results suggest that domain specialization is in itself not enough to guarantee global separability. Embeddings are not directly constrained geometries, so embeddings are not always fully disentangled across domains to be useful in real-world reviewer assignment settings with multiple and interdisciplinary research domains.



(a) Before contrastive learning

(b) After contrastive learning

Figure 2: Centroid Distance Distribution Before and After Contrastive Learning.

Table 2: Model-wise Similarity Statistics and Ranking Performance of Pretrained Embeddings.

Model	Domain	Cosine Similarity (Mean±Std)	MRR@10	NDCG@10
SPECTER	cs	0.101±0.004	0.3986	0.4118
	physics	0.067±0.006		
	math	0.011±0.003		
	q-bio	0.035±0.005		
	q-fin	0.029±0.004		
	stat	0.033±0.005		
	eess	-0.014±0.006		
BERT	cs	0.025±0.004	0.3715	0.3484
	physics	0.008±0.003		
	math	0.006±0.003		
BioBERT	math	0.004±0.002	0.3751	0.3549
MathBERT	physics	-0.001±0.003	0.3393	0.2691
FinBERT	math	0.008±0.003	0.3697	0.3428
PhysBERT	physics	0.026±0.004	0.3784	0.3668

Contrastive Geometry Refinement Results

We test the suggested Contrastive Geometry Refinement framework that explicitly reorganizes the embedding space by means of supervised contrastive optimization. The proposed method optimizes directly geometric properties to enhance separability and discriminability, as opposed to pretrained representations that depend on implicit semantic alignment.

The proposed framework achieves substantial improvements across all evaluation metrics, over pre-trained Specter Embedding model as shown in Table 3. MRR@10 increases from 0.3986 to 0.7715, and NDCG@10 improves from 0.4118 to 0.8289, indicating significantly better ranking quality. Top-1, Top-3, and Top-5 accuracies reach 0.6164, 0.9121, and 0.9737 respectively.

One of the results of contrastive optimization is the increased margins of similarity. The average similarity gap goes up to around 0.3332, which is significantly large, indicating that there is a significant rise in the separation of correct and competing assignments. This amplification of margins is the direct cause of the stability of ranking decisions and their confidence.

Also, we find that there are larger inter-domain centroid distances and smaller intra-domain variance which indicates that the learned representation is both smaller and more separable. These are geometric enhancements that contribute to the more distinct decision boundaries in the embedding space.

Notably, we find a causation between geometry and performance: the better the geometric separation, the bigger the margins, which consequently translate to more confident ranking and higher accuracy of retrieval. In contrast to pretrained embeddings, where the high similarity can be due to an overlapping distribution, the framework presented is such that the similarity is relative and absolute.

These results are also confirmed by centroid distance analysis. Contrary to expectation, contrastive learning, as demonstrated in Figure 2, vastly expands inter-domain distances and shrinks intra-domain spread resulting in well-separated and compact clusters.

Visualization and ROC Analysis

In order to examine the structural characteristics of the embedding space some more, we visualize representations with t-SNE. Geometrically, the t-SNE visualizations in Figure 3 indicate that pretrained embeddings have overlapping and entangled clusters, and contrastively refined embeddings have compact and well-separated clusters of domains. This restructuring is also confirmed through centroid-based analysis which indicates higher inter-domain centroid distances and smaller intra-domain variance.

Improved separability between correct and incorrect assignments is also further supported by ROC-AUC analysis as in Figure 4. The smoothed embeddings allow a more sound-based thresholding mechanism, meaning that the model allows confident and consistent decision-making in the context of beyond ranking-based evaluation.

DISCUSSION

The results of the experiment show that contrastive geometry refinement causes a radical reorganization of the embedding space with a trend towards steady improvement in ranking performance and assignment reliability. The proposed framework implements intra-domain compactness and inter-domain dispersion, which

makes the representation space consistent with the fundamental needs of matching reviewers and manuscripts.

Table 4 shows a domain-based analysis of the ranking performance. The findings indicate that all domains are improving in line with the proposed framework, meaning that the global geometric structure of the embedding space is improved and not that it is overfitting to certain categories.

Interestingly, domains like eess, stats, physics, etc. are the most successful, implying that those areas gain considerably on the compactness of clusters and the difference between domains. Conversely, other domains like the ones under cs and quantitative finance, have relatively lower scores that can also be explained by topical diversification and the presence of more semantic overlap and less distinct domain delimitation.

The overall enhancement of all spheres underscores how solid the approach is in terms of managing heterogeneous and interdisciplinary reviewer pools. This implies that when contrastive geometry refinement is applied, it does not refine the local similarity relationships but reorganizes the embedding manifold on a global scale.

Geometrically, the gained advantages can be attributed to a well-defined causal mechanism. Contrastive learning further widens inter-domain centroid gaps and narrows intra-domain variance resulting in higher separability of clusters. This reorganization has a direct positive impact on the similarity margins between correct and competing reviewers, leading to more stable ranking positions and assignment confidence. As a result, better rankings of metrics become a direct consequence of better geometric organization and not by-products of optimization.

In recent scholarly ecosystem, the volume of interdisciplinary domain manuscripts (e.g., Bioinformatics, computational neuroscience, optimization) are increasing. This makes the strict or hard domain separation in practical. The proposed framework deals with projection of the embedding space, which is having the continuous nature. This embedding space positions the manuscript based on the semantic similarity, allowing interdisciplinary manuscripts to reside in transition regions between domain clusters. These regions reflect shared semantic structure and are preserved during contrastive refinement.

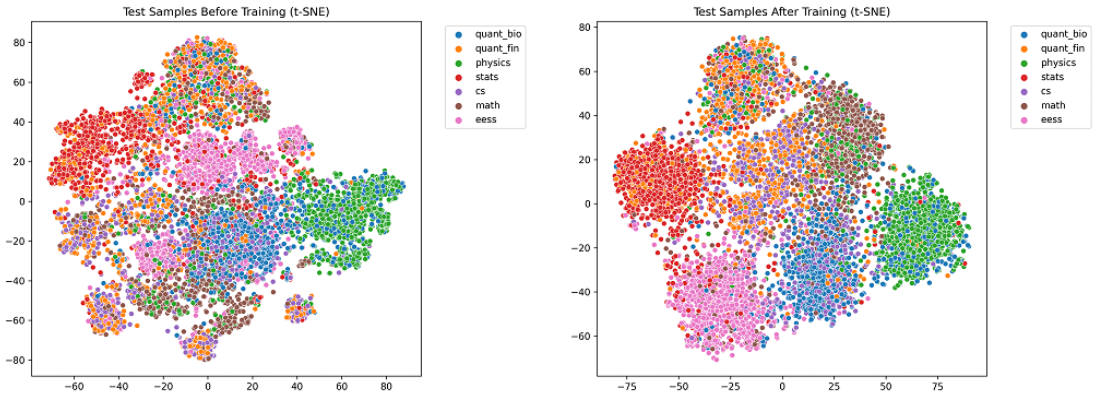
However, assignment confidence is explicitly modeled using margin-based and distribution-aware measures. For interdisciplinary manuscripts, similarity scores across competing reviewers tend to be closer, resulting in smaller margins and lower standardized confidence.

To further improve modeling of overlapping expertise, future extensions may incorporate multi-label contrastive learning, soft domain memberships, and hierarchical domain structures, enabling more flexible and realistic representation of scholarly domains.

Furthermore, the results highlight a key limitation of conventional similarity-based approaches. The pretrained embeddings tend to

Table 3: **Performance Comparison: Pretrained vs Contrastive Geometry Refinement.**

Model	Accuracy			Ranking		Geometry
	Top-1	Top-3	Top-5	MRR@10	NDCG@10	Avg Gap
SPECTER	0.32	0.68	0.85	0.3986	0.4118	0.07
Contrastive (Ours)	0.6164	0.9121	0.9737	0.7715	0.8289	0.3332



(a) Pretrained embeddings showing inter-domain overlap

(b) After contrastive refinement showing well-separated clusters

Figure 3: t-SNE Visualization of Embedding Space Geometry Before and After Contrastive Refinement.

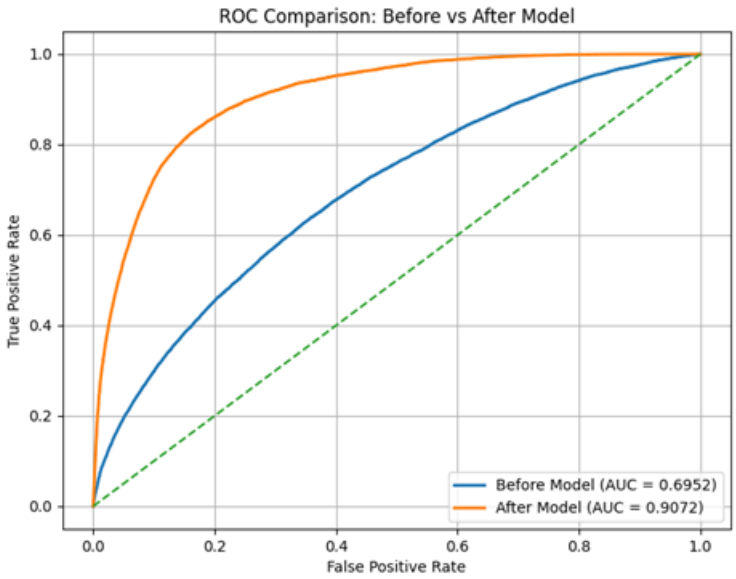


Figure 4: ROC-AUC curve illustrating improved discrimination after contrastive refinement. The Figure demonstrates enhanced separability between correct and incorrect assignments due to improved embedding geometry.

Table 4: Domain-wise Comparison of Ranking Performance Before and After Contrastive.

Category	Pretrained (Existing)		Contrastive Supervised (Proposed)	
	NDCG@10	MRR@10	NDCG@10	MRR@10
cs	0.3828	0.3560	0.7232	0.6272
quant_bio	0.3908	0.3677	0.7833	0.7098
math	0.4153	0.4038	0.8299	0.7738
physics	0.4214	0.4158	0.8445	0.7949
stats	0.4402	0.4405	0.8962	0.8622
quant_fin	0.3809	0.3532	0.8123	0.7473
eess	0.4513	0.4543	0.9106	0.8860
OVERALL	0.4118	0.3986	0.8289	0.7715

generate competitive ranking measures when they have similarity distributions that overlap, obscuring the underlying ambiguity of the assignment decisions. By contrast, the proposed framework will guarantee that similarity is relatively and absolutely discriminative and allows selecting reviewers more reliably and interpretably.

Overall, such results prove that the incorporation of geometry is one of the key predictors of the reliability of assignments. The proposed framework offers a principled way to enhance the performance of not only the ranking but also the confidence of a system of reviewer-manuscript matching by explicitly modeling and optimizing the geometric properties.

Although the suggested framework shows significant gains in implementing structure and prioritizing performance, a few shortcomings need to be explored. First, the framework is based on coarse domain-level supervision that might not sufficiently reflect fine-grained differences in expertise in sub-domains or very interdisciplinary research fields. Consequently, the learned embedding space might not best represent the reviewers with subtle or cross-domain knowledge. In Computer Vision, for instance, reviewers can be experts on databases but may have very different reviewing skills, and highly skilled in computer vision may have very different skills when it comes to database or machine learning. Likewise, semantically related domains like Machine Learning and Statistical Machine Learning can overlap, but have different reviewer communities. This intra-domain heterogeneity could help to create a simulated overlap in the embedding space, leading to high similarity matrix and low confidence margins during reviewer discrimination. However, it is still possible to apply the proposed framework in finer-grained

supervision schemes. Future research will explore the use of hierarchical representation of reviewers, which would combine multiple levels of expertise (at domain, sub-domain, and topic) represented through publication venues, citation networks, research taxonomies, and topic-aware representations. These extensions can enhance the geometric separability and precision for assignments in large-scale scholarly system for recommending reviewers in the reviewer recommendation system.

Second, the existing method uses fixed embeddings and does not consider the time dynamics, i.e., changing research interest, new topics, or shifts of the reviewer expertise with time. The system may be enhanced by adding a time or sequence modeling.

Additionally, the assessment is based on the embedding of quality and ranking-based measures, without considering real-world limitations like availability of reviewers, workload balancing, identifying of conflicts of interest, and considerations of fairness. These aspects are essential to practical implementation and need to be combined with system-level optimization systems.

These limitations will be critical to making geometry-conscious representation learning translate into entirely deployable reviewer assignment systems.

CONCLUSION

This paper re-examines reviewer-manuscript matching in a geometrical viewpoint, demonstrating that pretrained embeddings have inter-intra domain overlap and poor separability, resulting in inaccurate similarity-based assignment. To handle this, we proposed a Contrastive Geometry Refinement Framework which enforces compactness intra-domain and dispersion inter-domain by using supervised contrastive learning.

Theoretical and empirical findings prove that, better embedding geometry to achieve greater similarity gaps, decision boundaries, and much better ranking performance, with up to 2 times improvement in MRR and NDCG. These results prove the importance of embedding geometry in trusted reviewer assignment.

Future research directions involve generalizing the framework to fine-grained and hierarchical space, adaptive contrastive strategies (e.g., hard-negative mining), and metadata cues, e.g., citation and co-authorship networks. Also, it will be significant to model the temporal dynamics of the expertise of reviewers and consider practical constraints such as fairness and workload balancing to be applied in practice.

Additionally, the inclusion of hybrid supervision strategies, combining metadata (e.g., keywords, citations, venues) with weakly supervised or self-supervised signals, can further improve robustness incomplete labeling conditions. Together, these directions aim to make the framework more adaptive, scalable, and reliable for deployment in real-world scholarly ecosystems.

ABBREVIATIONS

MRR: Mean Reciprocal Rank; **NDCG:** Normalized Discounted Cumulative Gain; **BERT:** Bidirectional Encoder Representations from Transformers; **SciBERT:** Scientific Bidirectional Encoder Representations from Transformers; **BioBERT:** Biomedical Bidirectional Encoder Representations from Transformers; **MathBERT:** Mathematical Bidirectional Encoder Representations from Transformers; **PhysBERT:** Physics Bidirectional Encoder Representations from Transformers; **FinBERT:** Financial Bidirectional Encoder Representations from Transformers; **SPECTER:** Scientific Paper Embeddings using Citation-informed Transformers for Embedding Representations; **MIPS:** Maximum Inner Product Search; **ROC-AUC:** Receiver Operating Characteristic–Area Under the Curve; **t-SNE:** t-distributed Stochastic Neighbor Embedding; **TF-IDF:** Term Frequency–Inverse Document Frequency; **LSA:** Latent Semantic Analysis; **EESS:** Electrical Engineering and Systems Science; **CS:** Computer Science; **Q-Bio:** Quantitative Biology; **Q-Fin:** Quantitative Finance; **SupCon:** Supervised Contrastive Learning; **ReLU:** Rectified Linear Unit; **BatchNorm:** Batch Normalization; **Top-1:** Top-One Accuracy; **Top-3:** Top-Three Accuracy; **Top-5:** Top-Five Accuracy; **Std:** Standard Deviation; **Mann–Whitney U Test:** Nonparametric Statistical Test for Comparing Two Independent Samples; **Kolmogorov–Smirnov Test (KS Test):** Nonparametric Test for Comparing Two Probability Distributions; **K:** Number of Candidate Reviewers; **N:** Number of Manuscripts; **d:** Embedding Dimension; **M:** Manuscript Set; **R:** Reviewer Set; **MRR@10:** Mean Reciprocal Rank at Top 10; **NDCG@10:** Normalized Discounted Cumulative Gain at Top 10; **Top-K:** Top-K Ranked Candidates; **ROC:** Receiver Operating Characteristic; **AUC:** Area Under the Curve; **CS–Math:** Computer Science–Mathematics Cross-Domain Similarity; **Avg Gap:** Average Similarity Margin Gap; **Stats:** Statistics; **Quant_Bio:** Quantitative Biology; **Quant_Fin:** Quantitative Finance.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A Pretrained Language model for Scientific text. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 3613-3618. https://doi.org/10.18653/v1/d19-1371

Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., & Hullender, G. (2005). Learning to rank using gradient descent. In Proceedings of the 22nd International Conference on Machine Learning (ICML '05). Association for Computing Machinery, New York, NY, USA, 89-96. https://doi.org/10.1145/1102351.1102363

Charlin, L., & Zemel, R.S. (2013). The Toronto Paper Matching System: An automated paper-reviewer assignment system.

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In International Conference on Machine Learning (pp. 1597-1607). PMLR.

Cohan, A., Feldman, S., Beltagy, I., Downey, D., & Weld, D. S. (2020). SPECTER: Document-level representation learning using citation-informed transformers. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 2270-2282). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.586

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. Journal of the American Society for Information Science, 41(6), 391-407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171-4186). Association for Computational Linguistics.

Ethayarajh, K. (2019). How Contextual are Contextualized Word Representations? Comparing the Geometry of BERT, ELMo, and GPT-2 Embeddings. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 55-65, Hong Kong, China. Association for Computational Linguistics.

Goldsmith, J., & Sloan, R. H. (2007). The conference paper-chair assignment problem. Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data. https://doi.org/10.1145/1247480.1247555

Gu, Y., Dong, L., Wang, L., & Tresp, V. (2022). Domain-specific BERT models for scholarly document processing. Journal of Machine Learning Research.

Hellert, T., Montenegro, J., & Pollastro, A. (2024). PhysBERT: A text embedding model for physics scientific literature. APL Machine Learning, 2(4), 046105. https://doi.org/10.1063/5.0238090

Hogg, R. V., McKean, J. W., & Craig, A. T. (2013). Introduction to mathematical statistics (8th ed.). Pearson.

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. Pattern Recognition Letters, 31(8), 651-666.

Jolliffe, I. (2011). Principal component analysis. In: Lovric, M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_455

Kanwal, T., & Amjad, T. (2024). Research paper recommendation system based on multiple features from citation network. Scientometrics, 129(9), 5493-5531. https://doi.org/10.1007/s11192-024-05109-w

Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., & Krishnan, D. (2020). Supervised contrastive learning. In Proceedings of the 34th International Conference on Neural Information Processing Systems (Vol. 33, pp. 18661-18673). Curran Associates, Inc.

Kullmann, S., & Weimer, V. (2024). Developing a scientometric framework for open educational resources. Scientometrics, 129, 6065-6087. https://doi.org/10.1007/s11192-024-05007-1

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: A pre-trained biomedical language representation model for biomedical text mining. Bioinformatics. https://doi.org/10.1093/bioinformatics/btz682

Liu, Z., Huang, D., Huang, K., Li, Z., & Zhao, J. (2021). FinBERT: A pre-trained financial language representation model for financial text mining. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI'20.

Long, P. M., Jr., & Wong, J. (2013). On the complexity of matching problems. Journal of Artificial Intelligence Research.

Mimno, D., & McCallum, A. (2007). Expertise profiling for community-based collection development. In Proceedings of the 29th European Conference on IR Research (pp. 675-682). Springer. https://doi.org/10.1007/978-3-540-71496-5_64

Reusch, A. (2022). Pre-training for mathematics-aware retrieval. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (p. 3496). Association for Computing Machinery. https://doi.org/10.1145/3477495.3531680

Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. Communications of the ACM. https://doi.org/10.1145/361219.361220

Singh, C., Wang, F., & Li, Y. (2023). SciRepEval: A benchmark for scientific document representations. Transactions of the Association for Computational Linguistics, 11, 760-781. https://doi.org/10.1162/tacl_a_00573

Stelmakh, I., Wieting, J., Xi, S., Neubig, G., & Shah, N. B. (2023). A gold standard dataset for the reviewer assignment problem. arXiv. https://doi.org/10.48550/arXiv.2303.16750

Tang, L., Zhang, Y., Jin, R., Xia, H., & Zhang, B. (2012). A content-based recommendation system for peer reviewers. Proceedings of the 21st ACM International Conference on Information and Knowledge Management. https://doi.org/10.1145/2396761.2398454

Wang, F., Li, Y., & Zhang, J. (2023). exHarmony: Authorship and citations for benchmarking the real-world performance of reviewer recommendation. Proceedings of SIGIR.

Wang, F., Li, Y., & Zhang, J. (2024). Neural reviewer assignment with metadata. Proceedings of the ACM Web Conference. https://doi.org/10.1145/3589334.3648165

Wang, T., & Isola, P. (2020). Understanding contrastive representation learning through alignment and uniformity. ICML.

Zhang, F., & Wu, S. (2024). Predicting citation impact of academic papers across research areas using multiple models and early citations. Scientometrics, 129, 4137-4166. https://doi.org/10.1007/s11192-024-05086-0

Zhang, Y., Shen, Y., Kang, S. K., Chen, X., Jin, B., & Han, J. (2021). Neural reviewer assignment with metadata. Proceedings of the Web Conference.

Cite this article: Kotak N, Gohel B, Bavarva A. A Contrastive Supervised Learning Framework for Reliable Reviewer-Manuscript Matching. J Scientometric Res. 2026;15(2):386-97.